

ANGLIA RUSKIN UNIVERSITY

LORD ASHCROFT INTERNATIONAL BUSINESS SCHOOL

MOBILE INTERNET CONNECTIVITY,
EXPLORING STRUCTURAL BOTTLENECKS IN TAMIL NADU
USING ACTIVE INTERNET PERIPHERY MEASUREMENTS

SEBASTIAN SIGLOCH

A thesis in partial fulfilment of the
requirements of Anglia Ruskin University
for the degree of Doctor of Philosophy

Submitted: April 2018

ACKNOWLEDGEMENTS

It is my greatest pleasure to thank some of my advisors, acquaintances, companions, friends and peers that made this life-changing journey extremely joyful and rewarding.

First and foremost, I would like to express my greatest gratitude to my wonderful supervisory team, Prof Emanuele Giovannetti, fellow at Hughes Hall College at University of Cambridge and Professor of Economics Anglia Ruskin University and Dr Shailaja Fennell, Lecturer from the Centre of Development Studies at the University of Cambridge. Thank you for providing your consistent guidance, your valuable encouragement, and your constant support. You kept me aboard this ship, and I dearly hope that our professional paths continue to cross.

Moreover, I greatly express my gratitude to the Lord Ashcroft International Business School, for awarding me with a generous studentship that covered most of my expenses throughout a large part of this PhD journey. My greatest gratitude is also expressed to the Centre of Development Studies at the University of Cambridge, for covering the travel expenses for the primary data collection in Tamil Nadu, India, as part of an UKIERI research grant. I would also like to thank Prof Luciano Lenzini and Dr Valerio Luconi from the Istituto di Informatica e Telematica (IIT) at the University of Pisa. Their help in utilising the Portolan Network Sensing application was indisputably crucial for the outcome of this dissertation. In this sense, I would also like to thank the great Prof Ashok Jhunjhunwala and the entire Telecommunications and Computer Networks team from the Indian Institute of Technology Madras (IITM) for their support and hospitality during the data collection stopover in Chennai, Tamil Nadu, India. My thanks are also due to Dr Christine Lanner and Dr Andreas Dudler who thoroughly supported me on this journey during my work at the Swiss National Research and Education Network, SWITCH. Your outstanding work pioneers the Internet every day. I would also like to thank Hoi Lei Chan from the Anglia Ruskin University for her valuable support during the second half of my PhD journey. All of your motivational messages helped me keeping track. I would also like to thank my fellow research students: Dr Sandra Selmanovic, Dr Irina Popova, Dr Roman Eliassen and Dr Imko Meyenburg for joining me on numerous joyful philosophical discussions, especially throughout the first half of my journey.

My greatest thanks for her unbelievable mental support goes to Eunbi Koh. Thank you for always reminding me of my long-term visions and goals. Moreover, I would like to greatly thank Andreas Daiker, Tobias Ricken, Alex Anca Lauer, Athaulpa and other numerous not namely mentioned friends and acquaintances for their heartfelt friendship. Thank you for understanding my lack of time and focus. Warmly, I would also like to thank Anaïs Rousselle and Haika for their enduring and loving support along the biggest part of this journey. Please excuse my moments of prioritising work over you at the time. Quite unusually, I would also like to thank Volbeat, Trapt, Breaking Benjamin, I Prevail and Adelitas Way for their musical support during endless nights of writing ecstasies.

Lastly, and very heartfully, I would like to express my utmost gratitude to my parents, Alfred and Martina Sigloch, my living and deceased grand and great grandparents, my godmother Christa Hanke and the rest of my large and loving family. Great thanks are also due to my loving siblings Michel-Francois and Rebecca Sigloch. This work is for all of you. Without your everlasting support and understanding, I would have lost my sanity. Thank you all for sailing this ship with me. I dearly owe you!

ANGLIA RUSKIN UNIVERSITY

ABSTRACT

LORD ASHCROFT INTERNATIONAL BUSINESS SCHOOL

DOCTOR OF PHILOSOPHY

MOBILE INTERNET CONNECTIVITY,

EXPLORING STRUCTURAL BOTTLENECKS IN TAMIL NADU

USING ACTIVE INTERNET PERIPHERY MEASUREMENTS

SEBASTIAN SIGLOCH

April 2018

Accessible and affordable access to the Internet is crucial for socio-economic progress in developing countries and reducing Digital Divide. The disparity in mobile broadband penetration between urban and rural areas in the Indian state of Tamil Nadu can be explained by per capita income disparities. However, despite the vast body of multidisciplinary research, there are still large gaps in understanding Tamil Nadu's upstream Internet market structure and its impact on mobile broadband affordability. Moreover, there is a lack of research analysing the Internet market structure in developing countries using Network Analysis. This dissertation explores the presence of structural connectivity bottlenecks in the upstream Internet market for three mobile operator networks in Tamil Nadu. The exploration employs Complex and Statistical Network Analysis on primary data collected via active Internet periphery measurements through the Portolan application. The results obtained indicate the existence of hierarchical upstream Internet market structures for all operator networks. Moreover, the collected evidence indicates the reliance of mobile operator's connectivity on Tier-1 Internet Service Providers, while also revealing new Autonomous System relationships. This collected evidence highlights the crucial role that the level of hierarchical structuring of upstream Internet market structures plays in determining affordability. We show that end-users' prices per Megabyte increase with the level of hierarchical structuring, indicating the policy relevance of assessing Complex Network metrics to understand and address the hierarchical structuring of the relevant markets. In conclusions, this work indicates the importance of studying structural bottlenecks and connectivity hubs, as our evidence shows that the upstream Internet market structure also defines the bargaining powers exerted by Internet Service Providers, resulting in reduced competition and less affordable price plans. These results should also nudge policymakers' efforts to consider the different roles of 'bottlenecks' and 'hub-like' Internet Service Providers when aiming to reduce the Digital Divide.

Key words: Digital Divide, Mobile Internet Connectivity, Network Analysis, India

TABLE OF CONTENTS

1 Introduction.....	1
2 Literature Review.....	7
2.1 Upstream Interconnections	8
2.1.1 Nascent Stages of the Internet	8
2.1.2 Internet Service Provider Relationships	9
2.2 Mobile Broadband and Digital Divide in India	17
2.2.1 Indian Internet Infrastructure	17
2.2.2 Mobile Broadband in Tamil Nadu	20
2.2.3 The Next Big Frontier	30
2.2.4 Socio-Economic Impact of Mobile Broadband.....	32
2.3 Internet Market Structuring	39
2.3.1 Collection of Network Data.....	40
2.3.2 Network Formation	43
2.3.3 Network Structuring	47
2.3.4 Graph Visualisations and Layouts	53
2.4 Key Gaps in the Literature.....	60
2.5 Abduction.....	60
2.6 Conceptual Framework.....	61
2.7 Research Aims and Objectives.....	62
2.7.1 Research Aims	62
2.7.2 Research Objectives.....	63
3 Methodology	66
3.1 Pragmatism.....	66
3.2 Abductive Approach to Research	68
3.3 Exploratory Research based on a Single Case Study Strategy	69
3.3.1 Exploratory-Quantitative Research Design.....	69
3.3.2 Hypotheses-Generating Single Case Study Strategy	71
3.3.3 Cross-Sectional Study Design	72
3.3.4 Mobile Crowdsourced Primary Data Collection	73
3.4 Network Analytical Multimethod Research.....	77
3.4.1 Descriptive Statistics.....	79
3.4.2 Complex Network Analysis.....	79
3.4.3 Graph Visualisation Analysis	96
3.5 Exploratory Steps of Analysis.....	101

3.5.1 Iteration 1: Descriptive Analysis.....	101
3.5.2 Iteration 2: Complex Network Analysis (IP).....	103
3.5.3 Iteration 3: Graph Visualisation Analysis (IP)	105
3.5.4 Iteration 4: Complex Network Analysis (AS).....	108
3.5.5 Iteration 5: Graph Visualisation Analysis (AS)	110
3.5.6 Iteration 6: Autonomous System Relationships	111
3.5.7 Iteration 7: Statistical Network Analysis	115
3.6 Ethics, Biases, Reliability, Validity & Generalisability	117
3.6.1 Biases.....	117
3.7 Summary.....	118
4 Complex Network Analysis.....	120
4.1 Descriptive Network Analysis.....	121
4.2 Complex Network Analysis at IP granularity.....	125
4.2.1 General Network Metric Analysis (IP)	125
4.2.2 Edge Metric Analysis (IP).....	128
4.2.3 Vertex Metric Analysis (IP)	129
4.2.4 Complex Network Properties (IP).....	131
4.2.5 Summary Network Metric Analysis (IP)	133
4.2.6 Graph Visualisation Analysis (IP).....	134
4.2.7 Summary Complex Network Analysis (IP)	149
4.3 Complex Network Analysis at AS granularity.....	150
4.3.1 Data Preparations.....	150
4.3.2 General Network Metric Analysis (AS).....	152
4.3.3 Edge Metric Analysis (AS)	153
4.3.4 Vertex Metrics (AS).....	155
4.3.5 Complex Network Properties (AS).....	157
4.3.6 Summary Network Metric Analysis (AS)	160
4.3.7 Graph Visualisation analysis (AS)	160
4.3.8 Summary Complex Network Analysis (AS)	183
4.4 Autonomous System Relationships.....	184
4.4.1 Aircel Autonomous System Relationships.....	184
4.4.2 Bharti Airtel Autonomous System Relationships.....	189
4.4.3 Vodafone Autonomous System Relationships	193
4.4.4 Summary Autonomous System Relationships.....	198
4.5 Summary Complex Network Analysis.....	198
4.6 Key Findings Complex Network Analysis	199

5 Statistical Network Analysis.....	200
5.1 Review of Working Hypotheses	201
5.2 Background.....	202
5.3 Hierarchical Structuring	203
5.3.1 Descriptive Statistics.....	204
5.3.2 Econometric Models.....	224
5.3.3 Model 1 (Weighted Out-Degree connectivity).....	230
5.3.4 Model 2 (Weighted In-Degree)	237
5.3.5 Key Findings Hierarchical Structuring	245
5.4 Affordability and Quality of Service.....	246
5.4.1 Descriptive Statistics.....	248
5.4.2 Model 3.....	257
5.4.3 Model 4.....	263
5.4.4 Correlation with Quality of Service	266
5.4.5 Key Findings Affordability and Quality of Service.....	267
5.5 Key Findings Statistical Network Analysis.....	268
5.6 Summary Statistical Network Analysis.....	268
6 Discussion.....	269
6.1 Hierarchical Structuring	269
6.1.1 Hierarchical Upstream Internet Access Market.....	270
6.1.2 Structural Bottlenecks.....	274
6.1.3 Hidden Autonomous System Relationships.....	276
6.1.4 Summary Hierarchical Structuring.....	279
6.2 Hierarchical Structuring and Affordability.....	279
6.3 Hierarchical Structuring and Quality of Service.....	281
6.4 Implications to Practice and Policy.....	282
6.4.1 Implications to Practice.....	282
6.4.2 Implications to Policy	285
7 Contributions, Limitations, Generalisability and Recommendations	290
7.1 Contributions to Knowledge	290
7.2 Assessment of Working Hypotheses.....	292
7.3 Case Study Limitations	293
7.4 Generalisability and Transferability	295
7.5 Recommendations for Future Research	296
7.6 Conclusions.....	298
8 References.....	300

9 Appendices.....332

Ethical Considerations 346

Reliability..... 346

Pragmatic Validity..... 346

Generalisability..... 347

LIST OF TABLES

TABLE 2-1: INDIAN BROADBAND SUBSCRIBER BASE PER SEGMENT, SOURCE: TRAI (2016C).	20
TABLE 2-2: INDIAN TOTAL SUBSCRIBER BASE (WIRELINE AND WIRELESS) 2016 FOR TAMIL NADU MOBILE BROADBAND OPERATORS, SOURCE: TRAI (2016C).....	23
TABLE 2-3: INDIAN TOTAL WIRELESS SUBSCRIBER BASE 2016 FOR TAMIL NADU MOBILE BROADBAND OPERATORS, SOURCE: TRAI (2016C).	24
TABLE 2-4: INDIAN RURAL MARKET SHARE IN 2016 FOR TAMIL NADU MOBILE BROADBAND OPERATORS, SOURCE: TRAI (2016C).....	25
TABLE 2-5: TAMIL NADU MARKET SHARE 2016 FOR TAMIL NADU MOBILE BROADBAND OPERATORS, SOURCE: TRAI (2016C).....	25
TABLE 2-6: PER CAPITA INCOME AT CONSTANT PRICES TAMIL NADU VS ALL INDIA, SOURCE: TN-GOV-IN (2015).....	27
TABLE 2-7: TAMIL NADU MOBILE BROADBAND OPERATOR QUALITY OF SERVICE BENCHMARK, SOURCE: TRAI (2016B).	30
TABLE 3-1: OBSERVATIONS PER DATA COLLECTION DAY.	77
TABLE 3-2: EXAMPLE ADJACENCY MATRIX.....	85
TABLE 4-1: TRACEROUTE HOP OBSERVATIONS BY MOBILE BROADBAND OPERATOR.	123
TABLE 4-2: ROUND-TRIP-TIME BY MOBILE BROADBAND OPERATOR.	124
TABLE 4-3: SKIP DISTRIBUTION BY MOBILE BROADBAND OPERATOR.	125
TABLE 4-4: GENERAL METRICS BY MOBILE BROADBAND OPERATOR AT IP GRANULARITY.	128
TABLE 4-5: EDGE METRICS BY MOBILE BROADBAND OPERATOR AT IP GRANULARITY...	129
TABLE 4-6: VERTEX METRICS BY MOBILE BROADBAND OPERATOR AT IP GRANULARITY.	130
TABLE 4-7: COMPLEX NETWORK INDICATORS BY MOBILE BROADBAND OPERATOR AT IP	

GRANULARITY.....	132
TABLE 4-8: HIGHEST K-CORE VERTICES BY MOBILE BROADBAND OPERATOR AT IP GRANULARITY.....	148
TABLE 4-9: VERTICES AND EDGES AT IP AND AS GRANULARITY PER MOBILE BROADBAND OPERATOR.....	151
TABLE 4-10: GENERAL METRICS BY MOBILE BROADBAND OPERATOR AT AS GRANULARITY.	152
TABLE 4-11: EDGE METRICS BY MOBILE BROADBAND OPERATOR AT AS GRANULARITY.	154
TABLE 4-12: VERTEX METRICS BY MOBILE BROADBAND OPERATOR AT AS GRANULARITY.	155
TABLE 4-13: COMPLEX NETWORK INDICATORS BY MOBILE BROADBAND OPERATOR AT AS GRANULARITY.....	157
TABLE 4-14: COMPARISON OF COMPLEX NETWORK MODEL PROPERTIES, SOURCE: NEFEDOV (2013).....	158
TABLE 4-15: AIRCEL HIGHEST K-CORE VERTICES AT AS GRANULARITY.....	177
TABLE 4-16: BHARTI AIRTEL HIGHEST K-CORE VERTICES AT AS GRANULARITY.....	179
TABLE 4-17: VODAFONE HIGHEST K-CORE VERTICES AT AS GRANULARITY.....	181
TABLE 4-18: AIRCEL CUSTOMER CONES.....	186
TABLE 4-19: AIRCEL AS RELATIONSHIPS.....	187
TABLE 4-20: AIRCEL AS RELATIONSHIPS OVERVIEW.....	188
TABLE 4-21: BHARTI AIRTEL CUSTOMER CONE.....	190
TABLE 4-22: BHARTI AIRTEL AS RELATIONSHIPS.....	191
TABLE 4-23: BHARTI AIRTEL AS RELATIONSHIPS OVERVIEW.....	192
TABLE 4-24: VODAFONE CUSTOMER CONES.....	194

TABLE 4-25: VODAFONE AS RELATIONSHIPS.....	195
TABLE 4-26: VODAFONE AS RELATIONSHIPS OVERVIEW.....	196
TABLE 5-1: DESCRIPTIVE STATISTICS.....	205
TABLE 5-2: DESCRIPTIVE STATISTICS BY MOBILE BROADBAND OPERATOR.....	207
TABLE 5-3: HIGHEST AND LOWEST WEIGHTED IN-DEGREE AND WEIGHTED OUT-DEGREE MATRIX.....	208
TABLE 5-4: HIGHEST CLUSTERING COEFFICIENT OBSERVATIONS.....	214
TABLE 5-5: HIGHEST EIGENVECTOR CENTRALITY OBSERVATIONS.....	215
TABLE 5-6: DETAILED DESCRIPTIVE STATISTICS.....	215
TABLE 5-7: LN-TRANSFORMED DESCRIPTIVE STATISTICS.....	216
TABLE 5-8: LN-TRANSFORMED DETAILED DESCRIPTIVE STATISTICS.....	217
TABLE 5-9: LN-TRANSFORMED VARIABLE CORRELATION COEFFICIENTS.....	219
TABLE 5-10: AIRCEL LN-TRANSFORMED DESCRIPTIVE STATISTICS.....	223
TABLE 5-11: BHARTI AIRTEL LN-TRANSFORMED DESCRIPTIVE STATISTICS.....	223
TABLE 5-12: VODAFONE LN-TRANSFORMED DESCRIPTIVE STATISTICS.....	224
TABLE 5-13: FUNCTIONAL FORM SPECIFICATION MODEL 1.....	227
TABLE 5-14: FUNCTIONAL FORM SPECIFICATION MODEL 2.....	228
TABLE 5-15: AIRCEL MULTIPLE ROBUST LOG-LOG REGRESSION MODEL 1.....	231
TABLE 5-16: BHARTI AIRTEL MULTIPLE ROBUST LOG-LOG REGRESSION MODEL 1.....	233
TABLE 5-17: VODAFONE MULTIPLE ROBUST LOG-LOG REGRESSION MODEL 1.....	235
TABLE 5-18: AIRCEL MULTIPLE ROBUST LOG-LOG REGRESSION MODEL 2.....	239
TABLE 5-19: BHARTI AIRTEL MULTIPLE ROBUST LOG-LOG REGRESSION MODEL 2.....	241

TABLE 5-20: VODAFONE MULTIPLE ROBUST LOG-LOG REGRESSION MODEL 2.....	243
TABLE 5-21: PRICE PLAN EXAMPLES.	248
TABLE 5-22: DESCRIPTIVE STATISTICS MODEL 3 AND MODEL 4.....	249
TABLE 5-23: AIRCEL DESCRIPTIVE STATISTICS MODEL 3 AND MODEL 4.....	250
TABLE 5-24: BHARTI AIRTEL DESCRIPTIVE STATISTICS MODEL 3 AND MODEL 4.....	250
TABLE 5-25: VODAFONE DESCRIPTIVE STATISTICS MODEL 3 AND MODEL 4.	251
TABLE 5-26: CATEGORICAL COVARIATES DERIVED FROM MODEL 1 AND MODEL 2, USED AS PROXIES FOR HIERARCHICAL STRUCTURING PER MOBILE BROADBAND OPERATOR..	255
TABLE 5-27: OLS LOG-LOG REGRESSION MODEL 3.1.....	258
TABLE 5-28: OLS LOG-LOG REGRESSION MODEL 3.2.....	261
TABLE 5-29: MULTIPLE OLS LOG-LOG REGRESSION MODEL 4.....	264
TABLE 5-30: QUALITY OF SERVICE (QOS) CORRELATION TABLE.....	267
TABLE 9-1: AIRCEL PRICE PLANS 24TH JANUARY 2015, SOURCE: GSMOUTLOOK (2015A).	340
TABLE 9-2: BHARTI AIRTEL PRICE PLANS 24TH JANUARY 2015, SOURCE: GSMOUTLOOK (2015B).....	341
TABLE 9-3: BSNL PRICE PLANS 19TH JANUARY 2015, SOURCE: GSMOUTLOOK (2015C).	341
TABLE 9-4: VODAFONE PRICE PLANS 25TH JANUARY 2015, SOURCE: GSMOUTLOOK (2015D).....	342
TABLE 9-5: KEY CHARACTERISTICS OF CASE STUDIES, SOURCE: STAKE (1995), YIN (2003), FLYVBJERG (2006) AND BRYMAN (2012).	343
TABLE 9-6: NUMBER OF OBSERVATIONS BY ORIGINATING AUTONOMOUS SYSTEM.	349
TABLE 9-7: AIRCEL DEGREE DISTRIBUTION USING R (2016).....	350

TABLE 9-8: AIRCEL K-CORE DECOMPOSITION USING R (2016).....	352
TABLE 9-9: AUTONOMOUS SYSTEM NUMBERS AND ORGANISATIONAL NAMES, SOURCE: MAXMIND (2015).....	353

LIST OF FIGURES

FIGURE 2-1: INTERNET SERVICE PROVIDER (ISP) CUSTOMER CONES.....	14
FIGURE 2-2: INTERNET PROTOCOL (IP) TIER NETWORKS.....	16
FIGURE 2-3: ARPANET 1969 GRAPH VISUALISATION, SOURCE: THE OCP (2016).....	53
FIGURE 2-4: RANDOM LAYOUT GRAPH VISUALISATION WITH 200 VERTICES AND 1333 EDGES, ELABORATED USING GEPHI (2016).....	55
FIGURE 2-5: LAYERED LAYOUT GRAPH VISUALISATION WITH 200 VERTICES AND 1333 EDGES, ELABORATED USING GEPHI (2016).....	56
FIGURE 2-6: FRUCHTERMAN-REINGOLD LAYOUT GRAPH VISUALISATION WITH 200 VERTICES AND 1333 EDGES, ELABORATED USING GEPHI (2016).....	57
FIGURE 2-7: FORCE ATLAS 2 LAYOUT GRAPH VISUALISATION WITH 200 VERTICES AND 1333 EDGES, ELABORATED USING GEPHI (2016).....	58
FIGURE 2-8: K-CORE DECOMPOSITION GRAPH VISUALISATION WITH 200 VERTICES AND 1333 EDGES, ELABORATED USING R (2016).....	59
FIGURE 2-9: CONCEPTUAL FRAMEWORK.....	62
FIGURE 3-1: OVERVIEW OF DATA COLLECTION.....	74
FIGURE 3-2: TRACEROUTE HOP OBSERVATIONS AS OBTAINED THROUGH PORTOLAN (2015) PLOTTED ON A GOOGLE MAPS.....	76
FIGURE 3-3: WARRANTED ASSERTIONS PROCESS TO INQUIRY, SOURCE: DEWEY (1941). .	78
FIGURE 3-4: EXAMPLE VERTICES AND EDGES.	81
FIGURE 3-5: EXAMPLE NETWORK DISPLAYING NETWORK DENSITY.....	84
FIGURE 3-6: EXAMPLE NETWORK DISPLAYING AN ADJACENCY MATRIX.....	86
FIGURE 3-7: EXAMPLE NETWORK DISPLAYING EDGE BETWEENNESS.	88
FIGURE 3-8: EXAMPLE NETWORK DISPLAYING NEIGHBOURHOOD OVERLAP.	89

FIGURE 3-9: EXAMPLE NETWORK DISPLAYING EMBEDDEDNESS.	91
FIGURE 3-10: SMALL-WORLD NETWORK GRAPH VISUALISATION.	98
FIGURE 3-11: SCALE-FREE NETWORK GRAPH VISUALISATION.	99
FIGURE 3-12: K-CORE DECOMPOSITION GRAPH VISUALISATION.	100
FIGURE 4-1: EXAMPLE OF COLLECTED PARIS TRACEROUTE OBSERVATIONS.	121
FIGURE 4-2: VERTEX STRENGTH DISTRIBUTIONS BY MOBILE BROADBAND OPERATOR AT IP GRANULARITY.	131
FIGURE 4-3: DEGREE DISTRIBUTIONS BY MOBILE BROADBAND OPERATOR AT IP GRANULARITY.	133
FIGURE 4-4: SMALL-WORLD NETWORK GRAPH VISUALISATIONS IN LAYERED LAYOUT BY MOBILE BROADBAND OPERATOR AT IP GRANULARITY.	136
FIGURE 4-5: BARABÁSI-ALBERT STANDARD MODEL GRAPH VISUALISATIONS PER MOBILE BROADBAND OPERATOR AT IP GRANULARITY.	139
FIGURE 4-6: BARABÁSI-ALBERT MODEL A GRAPH VISUALISATIONS PER MOBILE BROADBAND OPERATOR AT IP GRANULARITY.	141
FIGURE 4-7: BARABÁSI-ALBERT MODEL B GRAPH VISUALISATIONS PER MOBILE BROADBAND OPERATOR AT IP GRANULARITY.	143
FIGURE 4-8: AIRCEL GRAPH VISUALISATION K-CORE DECOMPOSITION AT IP GRANULARITY.	145
FIGURE 4-9: BHARTI AIRTEL GRAPH VISUALISATION K-CORE DECOMPOSITION AT IP GRANULARITY.	146
FIGURE 4-10: VODAFONE GRAPH VISUALISATION K-CORE DECOMPOSITION AT IP GRANULARITY.	147
FIGURE 4-11: VERTEX STRENGTH DISTRIBUTION BY MOBILE BROADBAND OPERATOR AT AS GRANULARITY.	156
FIGURE 4-12: DEGREE DISTRIBUTIONS BY MOBILE BROADBAND OPERATOR AT AS	

GRANULARITY.	159
FIGURE 4-13: SMALL-WORLD NETWORK GRAPH VISUALISATIONS IN LAYERED LAYOUT BY MOBILE BROADBAND OPERATOR AT AS GRANULARITY.	162
FIGURE 4-14: BARABÁSI-ALBERT MODEL B GRAPH VISUALISATIONS PER MOBILE BROADBAND OPERATOR AT AS GRANULARITY.....	165
FIGURE 4-15: AIRCEL GRAPH VISUALISATIONS CENTRALITY METRICS USING THE FRUCHTERMAN-REINGOLD LAYOUT AT AS GRANULARITY.	170
FIGURE 4-16: BHARTI AIRTEL GRAPH VISUALISATIONS CENTRALITY METRICS USING THE FRUCHTERMAN-REINGOLD LAYOUT AT AS GRANULARITY.	172
FIGURE 4-17: VODAFONE GRAPH VISUALISATIONS CENTRALITY MEASUREMENTS USING THE FRUCHTERMAN-REINGOLD LAYOUT AT AS GRANULARITY.	174
FIGURE 4-18: AIRCEL GRAPH VISUALISATION K-CORE DECOMPOSITION AT AS GRANULARITY.	178
FIGURE 4-19: BHARTI AIRTEL GRAPH VISUALISATION K-CORE DECOMPOSITION AT AS GRANULARITY.	180
FIGURE 4-20: VODAFONE GRAPH VISUALISATION K-CORE DECOMPOSITION AT AS GRANULARITY.	182
FIGURE 4-21: AIRCEL GRAPH VISUALISATION WITH RELATIONSHIP COLOURING.....	189
FIGURE 4-22: BHARTI AIRTEL GRAPH VISUALISATION WITH RELATIONSHIP COLOURING.	193
FIGURE 4-23: VODAFONE GRAPH VISUALISATION WITH RELATIONSHIP COLOURING.	197
FIGURE 5-1: OVERVIEW OF ECONOMETRIC MODELS.	201
FIGURE 5-2: IN-DEGREE AND OUT-DEGREE FLOW THROUGH A VERTEX.....	206
FIGURE 5-3: TOTAL OBSERVATIONS SCATTER PLOT WEIGHTED IN-DEGREE AGAINST WEIGHTED OUT-DEGREE.....	209
FIGURE 5-4: AIRCEL SCATTER PLOT WEIGHTED IN-DEGREE AGAINST WEIGHTED OUT-	

DEGREE.	210
FIGURE 5-5: BHARTI AIRTEL SCATTER PLOT WEIGHTED IN-DEGREE AGAINST WEIGHTED OUT-DEGREE.	210
FIGURE 5-6: VODAFONE SCATTER PLOT WEIGHTED IN-DEGREE AGAINST WEIGHTED OUT-DEGREE.	211
FIGURE 5-7: CLUSTERING COEFFICIENT DISTRIBUTION WITH FIT LINE.	213
FIGURE 5-8: LN(WEIGHTED OUT-DEGREE) DISTRIBUTION WITH FIT LINE.	217
FIGURE 5-9: LN(WEIGHTED IN-DEGREE) DISTRIBUTION WITH FIT LINE.....	218
FIGURE 5-10: LN(CLUSTERING COEFFICIENT) DISTRIBUTION WITH FIT LINE.	218
FIGURE 5-11: LN(EIGENVECTOR CENTRALITY) DISTRIBUTION WITH FIT LINE.	219
FIGURE 5-12: TWO-WAY SCATTER PLOTS LN(CLUSTERING COEFFICIENT) AGAINST LN(WEIGHTED OUT-DEGREE) PER OPERATOR WITH LINEAR FIT LINE.....	220
FIGURE 5-13: TWO-WAY SCATTER PLOTS LN(EIGENVECTOR CENTRALITY) AGAINST LN(WEIGHTED OUT-DEGREE) PER OPERATOR WITH LINEAR FIT LINE.....	221
FIGURE 5-14: TWO-WAY SCATTER PLOTS LN(CLUSTERING COEFFICIENT) AGAINST LN(WEIGHTED IN-DEGREE) PER OPERATOR WITH LINEAR FIT LINE.	221
FIGURE 5-15: TWO-WAY SCATTER PLOTS LN(EIGENVECTOR CENTRALITY) AGAINST LN(WEIGHTED IN-DEGREE) PER OPERATOR WITH LINEAR FIT LINE.	222
FIGURE 5-16: AIRCEL PREDICTED ERROR TERM RESIDUAL DISTRIBUTION MODEL 1.	232
FIGURE 5-17: BHARTI AIRTEL PREDICTED ERROR TERM RESIDUAL DISTRIBUTION MODEL 1.	234
FIGURE 5-18: VODAFONE PREDICTED ERROR TERM RESIDUAL DISTRIBUTION MODEL 1.	236
FIGURE 5-19: PREDICTED EFFECTS OF AN INCREASE IN LN(CLUSTERING COEFFICIENT) IN LN(WEIGHTED OUT-DEGREE) - MODEL 1.	236
FIGURE 5-20: PREDICTED EFFECTS OF AN INCREASE IN LN(EIGENVECTOR CENTRALITY) IN	

LN(WEIGHTED OUT-DEGREE) - MODEL 1.....	237
FIGURE 5-21: AIRCEL PREDICTED ERROR TERM RESIDUAL DISTRIBUTION MODEL 2.....	240
FIGURE 5-22: BHARTI AIRTEL PREDICTED ERROR TERM RESIDUAL DISTRIBUTION MODEL 2.	242
FIGURE 5-23: VODAFONE PREDICTED ERROR TERM RESIDUAL DISTRIBUTION MODEL 2.	244
FIGURE 5-24: PREDICTED EFFECTS OF AN INCREASE IN LN(CLUSTERING COEFFICIENT) ON LN(WEIGHTED IN-DEGREE) - MODEL 2.	244
FIGURE 5-25: PREDICTED EFFECTS OF AN INCREASE IN LN(EIGENVECTOR CENTRALITY) ON LN(WEIGHTED IN-DEGREE) - MODEL 2.	245
FIGURE 5-26: PRICE DISTRIBUTION BY MOBILE BROADBAND OPERATOR.	251
FIGURE 5-27: DATA IN MEGABYTE DISTRIBUTION BY MOBILE BROADBAND OPERATOR.	252
FIGURE 5-28: VALIDITY IN DAYS DISTRIBUTION BY MOBILE BROADBAND OPERATOR.	252
FIGURE 5-29: TWO-WAY SCATTER PLOTS DATA IN MEGABYTE AGAINST PRICE PER MOBILE BROADBAND OPERATOR WITH LINEAR FIT LINE.	253
FIGURE 5-30: PRICE PER MEGABYTE DISTRIBUTION BY MOBILE BROADBAND OPERATOR.	254
FIGURE 5-31: INCREASE IN THE LCLUS_MODEL1_HAT PROXY.....	259
FIGURE 5-32: PREDICTED ERROR TERM RESIDUAL DISTRIBUTION MODEL 3.1.....	260
FIGURE 5-33: PREDICTED ERROR TERM RESIDUAL DISTRIBUTION MODEL 3.2.....	262
FIGURE 5-34: PREDICTED ERROR TERM RESIDUAL DISTRIBUTION MODEL 4.	265
FIGURE 9-1: AIRCEL AS55831 ROUTE PROPAGATION, SOURCE: HURRICANE ELECTRIC (2016).....	354
FIGURE 9-2: BHARTI AIRTEL AS45609 ROUTE PROPAGATION, SOURCE: HURRICANE ELECTRIC (2016).	354
FIGURE 9-3: VODAFONE AS38266 ROUTE PROPAGATION, SOURCE: HURRICANE ELECTRIC	

(2016).....	355
FIGURE 9-4: BHARTI AIRTEL AS9498 ROUTE PROPAGATION, SOURCE: HURRICANE ELECTRIC (2016).....	355
FIGURE 9-5: CHINA EDUCATION AND RESEARCH NETWORK CENTER AS4538 ROUTE PROPAGATION, SOURCE: HURRICANE ELECTRIC (2016).....	356
FIGURE 9-6: CABLE AND WIRELESS WORLDWIDE PLC. AS1273 ROUTE PROPAGATION, SOURCE: HURRICANE ELECTRIC (2016).	356
FIGURE 9-7: TATA COMMUNICATIONS (FORMERLY VSNL) AS4755 ROUTE PROPAGATION, SOURCE: HURRICANE ELECTRIC (2016).	357
FIGURE 9-8: 'THE EYE' - TOTAL 731,200 OBSERVATIONS AT IP GRANULARITY USING THE BARÁBASI-ALBERT STANDARD MODEL WITH THICK EDGES, ELABORATED USING GEPHI (2016).	376

LIST OF APPENDICES

PUBLICATIONS, CONFERENCE PRESENTATIONS AND ABSTRACTS	333
TAMIL NADU MOBILE BROADBAND OPERATOR PRICE PLANS.....	339
KEY CHARACTERISTICS OF CASE STUDIES	343
DATA COLLECTION INSTRUCTIONS	344
ETHICAL CONSIDERATIONS, RELIABILITY, VALIDITY AND GENERALISABILITY.....	346
NUMBER OF OBSERVATIONS BY ORIGINATING AUTONOMOUS SYSTEM	348
POWER-LAW DEGREE DISTRIBUTIONS USING R	350
K-CORE DECOMPOSITION USING R	351
AS NUMBERS AND ORGANISATIONAL NAMES	353
AUTONOMOUS SYSTEM ROUTE PROPAGATIONS	354
STATA DO-FILES	358
ADDITIONAL GRAPH VISUALISATIONS	376

TABLE OF ABBREVIATIONS AND ACRONYMS

Abbreviation	Description
ADI	Affordability Drivers Index
AIX	Amsterdam Internet Exchange
APNIC	Asia Pacific Network Information Center
ARPANET	Advanced Research Projects Agency Network
AS	Autonomous System
ASN	Autonomous System Number
AUP	Acceptable Use Policy
BGP	Border Gateway Protocol
BSNL	Bharat Sanchar Nigam Limited
CAIDA	Center for Applied Internet Data Analysis
CCS	Customer Cone Size
CDMA	Code-division multiple access
CDN	Content Delivery Network
CERFnet	California Education and Research Federation Network
CIDR	Class Interdomain Routing
CSNET	Computer Science Network
DARPA	Defense Advanced Research Projects Agency
DAG	Directed Acyclic Graph, a graph visualisation layout.
DIMES	EVERGROW Integrated Project in the EU Information Society Technologies, Future and Emerging Technologies programme.
DNS	Domain Name System
EFF	Electronic Frontier Foundation
ERNET	Educational Research Network
FCC	Federal Communication Commission of the United States of America.
FUP	Fair Usage Policy

GB	Gigabyte
GDI	Global Diffusion of the Internet
GDP	Gross Domestic Product
GIAS	Gateway Internet Access Service
GNI	Gross National Income
GSDP	Gross State Domestic Product
GSM	Global System for Mobile Communications
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IANA	Internet Assigned Numbers Authority
IAMAI	Internet and Mobile Association of India
IBEF	Indian Brand Equity Foundation
ICSI	Berkeley International Computer Science Institute
ICT	Information and Communication Technology
ICT4D	Information and Communication Technology for Development
IDI	International Development Index
IIT	Intituto di Informatica e Telematica of the Italian National Research Council CNR
IITM	Indian Institute of Technology Madras (Chennai)
Inc	Incorporated
INR	Indian Rupees
iOS	Mobile Operating System by Apple
IP	Internet Protocol
IPv4	Internet Protocol version 4
IRR	Internet Routing Registry
ISP	Internet Service Provider
ITU	International Telecommunications Union

IXP	Internet Exchange Point
kbps	Kilobyte per second
LINX	London Internet Exchange
MB	Megabyte
MFENet	Magnetic Fusion Energy Network of the United States Department of Energy
MIT	Massachusetts Institute of Technology
NASA	National Aeronautics and Space Administration
NREN	National Research and Education Network
NRI	Network Readiness Index
NSF	National Science Foundation
NSFNET	National Science Foundation Network
OECD	Organisation for Economic Cooperation and Development
OIO	Open Internet Order
OLS	Ordinary Least Squares
PDP	Packet Data Protocol
p.c.	Per capita
plc	Public limited company
PSINet	Performance Systems International Network
QoS	Quality of Service
RIR	Regional Internet Registry
RIPE NCC	Réseaux IP Européens Network Coordination Centre
RRC	Remote Route Collector
SDG	Sustainable Development Goals
SLAC	Stanford Linear Accelerator Center
TCP/IP	Transmission Control Protocol / Internet Protocol
TRAI	Telecom Regulatory Authority of India

UC	University of California
UKIERI	UK – India Education and Research Initiative
UN	United Nations
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
USB	Universal Serial Bus
VSAT	Very Small Aperture Terminal
VSNL	Videsh Sanchar Nigam Limited
WDR	World Development Report
WEF	World Economic Forum
WH	Working Hypothesis
WWW	World Wide Web

GLOSSARY

Term	Description
2G	2 nd Generation of Mobile Connectivity.
3G	3 rd Generation of Mobile Connectivity.
4G	4 th Generation of Mobile Connectivity.
Autonomous System (AS)	A connected group of one or more IP prefixes run by one or more network operators which has a singly and clearly defined routing policy (Hawkinson and Bates, 1996, p.2).
Backbone	The backbone of the Internet refers to key data routes between large, strategically important networks, composing the Internet, mostly through participation of Tier-1 <i>Internet Service Providers</i> .
Broadband	Connections with a download speed ≥ 512 Kbps (TRAI, 2016c, p.36).
Crore	Hindi numbering system denotation for 10,000,000 (ten million).
Fixed Wireless	A wireless connection through Wi-Fi, Wi-Max, Point-to-Point Radio & VSAT (see TRAI, 2016c, p.xii).
Internet Service Provider (ISP)	Is an organisation that provides services to access and use the Internet. An ISP can be an Access Provider, a Transit Provider, a Content Provider, a Content Distribution Network (CDN) or an Internet Exchange Point (IXP).
IP(v4) address	Unique identifier (in 32bit format) being assigned to devices (e.g. routers, mobile devices, computers) in a TCP/IP network (IPv4 or IPv6, dependent on the version adopted (in this dissertation IPv4)).
IP (address) prefix	IPv4 (Version 4) prefixes are patterns that match the first n binary bits of an IPv4 address. If an example IP address 128.8.0.0/16 can be written in 128.8/16. This means that the prefix matches 10000000 00001000 as the first sixteen bits. This would match e.g. an IPv4 address of 128.8.74.1, 128.8.8.8, or 128.8.0.0, but not 128.9.7.3. Every Autonomous System manages a multitude of these IP address prefixes, being able to uniquely identify a multitude of networking devices.
Mobile Wireless	A wireless connection through Phones or Dongles (see TRAI, 2016c, p.xii).
Narrowband	Connections with a download speed < 512 Kbps (TRAI, 2016c, p.38).

Packet Switching Networks	Refers to network communication, where transmitted data is separated into equally large packets to-be transferred from a data sender to a receiver using multiple connections.
Panyachats	Hindi word for a village council in India.
Peering	Voluntary interconnection, through Border Gateway Protocol (BGP) routing, amongst networks on the Internet that are governed by the administrative control of an Autonomous System. This interconnection usually refers to the mutual exchange of data under settlement-free (unpaid) arrangements. However, paid peering relationships may similarly exist.
Ping	Ping (or ping-time) refers to a computer network software to test the (time to) reach ‘hosts’ of an Internet Protocol (IP) network.
Round-Trip-Time (RTT)	Indicates the time a data packet takes to be sent from the initial source IP address to the destination one, plus the time it takes for this to be acknowledged by the destination IP address and returned to the source IP address.
Traceroute	Refers to a network diagnostics and measurement tool using the Paris Traceroute (2016) version. Traceroutes are used to display and measure the path from a source to a destination of data packets across Internet Protocol networks. The measurements are recorded as the Round-Trip-Time (RTT) that a data packet needs to being acknowledged by the destination. Each traceroute contains a number of hops (steps) along its given path to a destination. The sum of the mean time in each hop (traceroute step) measures the total time spent to establish a connection (RTT).
Transit	Refers to a service that allows network traffic on the Internet that allows the smaller Internet Service Providers to transit other networks, at a given transit cost, to reach other parts of the Internet.
Wi-Fi	Technology for Wireless local area networking with devices based on the IEEE 802.11 standards.
Wi-Max	Technology for Worldwide Interoperability for Microwave Access for wireless networking based on the IEEE 802.16 standards.
Wireless	Fixed and mobile wireless connections.

COPYRIGHT DECLARATION

I hereby declare that the work in this dissertation is my own. It contains no material that has been previously submitted, in part or whole, to any university or institution, for any award of a degree, diploma or other qualification. All use of sources is acknowledged in accordance with established rules on fair dealing for the purposes of review, criticism, referral, quotation and sound referencing practices.

In accordance with the Lord Ashcroft International Business School guidelines, this thesis does not exceed 80,000 words.

The copyright of this dissertation lies with

- i. Anglia Ruskin University, for one year starting from the date of publication, after which it is handed over to
- ii. Sebastian Sigloch

This copy of the dissertation has been supplied on the condition that anyone who consults it is bound by copyright.

04.04.2018

Sebastian Sigloch

Cambridge, United Kingdom

To my parents.

1 INTRODUCTION

'I do not fear the Internet. I fear its understatement and unequal access.'
(Sigloch, 2017).

The Internet is a collection of interconnecting networks operated by different types of network operators among which *Internet Service Providers (ISPs)* provide global connectivity access to final end-users. *ISPs* play a vital technical and economic role in shaping the global connectivity costs, underlying the Internet. Recent reports show that an equal and affordable access to the Internet has the power to induce substantial economic growth (GSMA, 2016; TRAI, 2016a), education and employment (Fennell et. al., 2016), inclusion (Broadband Commission, 2016), equality and social impact (WDR, 2016). The World Economic Forum rates affordable connectivity as a key infrastructural element for a robust digital economy (WEF 2017, p.11), providing global economies with the power to lift billions of people out of poverty (ITU, 2014). These benefits arise from an increased access to information, markets, and innovation possibilities, gained by end-users through their ability to connect to the Internet. Nevertheless, affordable and equal access to the Internet remains a key challenge to be solved (Broadband Commission, 2014; World Bank, 2016c). Especially developing and emerging economies, where most of the world's offline population resides (ITU, 2016), face those challenges, partly due to large International connectivity costs. Moreover, mobile broadband access to the Internet is highly significant for the accomplishment of the Sustainable Development Goals, adopted by the United Nations (2017). Recent data show that India ranks world second, behind China, for the absolute number of Internet users (Internet Live Stats, 2016). Nevertheless, India still shows a low 'Internet penetration rate' (ITU, 2015c).

With more than a billion subscribers, according to the latest available data from February 2016 (Gov-IN, 2017), mobile telephony is experiencing an unprecedented growth in India. Given the fast-track advancements of the country's mobile broadband infrastructure, most Indians are accessing the Internet through wireless (narrow and broadband) rather than wired technologies (TRAI, 2016c). However, only about 26 per cent of the Indian population had access to the Internet by the end of 2015 (ITU, 2015d). Ericsson (2016) and Statista (2016b) expect the number of Indian mobile broadband users to continue to grow significantly in the upcoming years, reaching more than 1.3 billion mobile cellular subscribers by 2021 (Ericsson, 2016, p.2), including 0.8 billion mobile broadband ones. Due to such predicted growth rates, India is considered to be the next big frontier in the digital world (Broadband Commission, 2016, p.19), providing India with unprecedented opportunities for socio-economic impact, yet also potential infrastructure access problems and bandwidth overloads. A careful management of the future digital expansion is therefore essential in facilitating India's digital development efforts. Besides mobile broadband providing an unparalleled opportunity for India, the affordability of mobile broadband price plans to access the global Internet infrastructure, especially in more rural areas, remain under-researched and yet present a very important issue to nurture access growth. By focussing on penetration at a more disaggregate, state, level, it is possible to capture interesting differences among the Indian states. Tamil Nadu is one of the states with the highest discrepancies between urban and rural mobile broadband subscribers (TRAI, 2016e; TRAI, 2017, p.16). Moreover, this state's districts show significant per capita income disparities (Sundar, 2015; Selvabaskar et al., 2016), with a large fraction of the population's income remaining below the World Bank poverty line (TN-GOV-IN, 2015). Hence, Tamil Nadu faces a relevant digital gap, driven by urban-rural income and gender disparities (TN-GOV-IN, 2015). These affordability disparities could be resulting from the fact that Indian mobile broadband operators have different structural properties in their upstream Internet access market, where these properties are the emerging macro-features, based on the micro, bilateral connectivity decisions shaped by the economic relationships between the local *Internet Service Providers* and their global, often much larger, counterparts. This abducted insight provides a key focus of this dissertation (see section 2.5 below).

Given these considerations, we argue that end-users' mobile broadband affordability is a critical issue for developing countries, given the underlying International connectivity costs for mobile broadband operators to access the global Internet infrastructure, as

further discussed in the literature review in Chapter 2.

Internet transit prices in the U.S. have been dropping consistently, as e.g. Norton (2010) points out. However, Internet end-users located in the Asia-Pacific region interestingly pay higher charges for International Internet connections than their global counterparts (Sultana, 2016). One reason for this is, that developing countries that aim to connect to the Internet backbone pay the full costs of international leased lines to the country, as Sultana (2016, p.26) points out at an ITU workshop in New Delhi. Such costs to access the upstream Internet market must then be passed on to the end-users, reducing affordability and thereby inhibiting mobile broadband adoptions.

India's recent low rankings in the Web Index of the World Wide Web Foundation (2014) or the Affordability Drivers Index (A4AI, 2016) might be outcomes from those high International Internet connectivity costs. Research that addresses structural upstream bottlenecks is still often neglected in the field of Internet Economics. This dissertation aims to utilise a combination of exploratory- and quantitative approaches to study the upstream Internet infrastructure, which might provide insights and methodological advancements into these issues. Ultimately, this work aims to help to prevent the widening of the present Indian, and in more detail, Tamil Nadu's Digital Divide (Sundar, 2015; WDR, 2016).

Local mobile broadband operators interconnect with larger regional or International *Internet Service Providers (ISPs)* through paid contracts or unpaid peering relationships to access the upstream Internet market. This is an essential precondition for the mobile broadband operators to acquire global end-to-end connectivity and service provisioning. Larger *Internet Service Providers* then route the traffic, received from the local operators, through the global digital supply chain, making use of registered Autonomous Systems to manage interoperator connectivity. The collection of relationship data between *Internet Service Providers*, underlying this internetworking process, has always been expensive and arduous (Newman, Barabási and Watts, 2006; Schneider and Bauer, 2016). The main resources and insights, about these relationships, came from the pioneering measurements work of the Center for Applied Internet Data Analysis (CAIDA) and the, more commercially oriented, reports by Telegeography. However, recent computing advancements are now enabling researchers to collect primary Internet connectivity data using crowdsourced active Internet periphery measurements (Faggiani et al., 2013), as used in this dissertation.

The structural features of the Internet, emerging from these connectivity data, are amenable to be studied using Complex Network Analysis, which has been, until recently, also an under-researched field (Gorman and Malecki, 2000). Vázquez, Pastor-Satorras and Vespignani (2002), for example, study the structure of the Internet by analysing the connectivity of 6,374 Autonomous Systems using Complex Network Analysis. Similarly, Barnett and Park (2012) investigate the network structure of websites on the World Wide Web. More closely related to this dissertation, Ruiz and Barnett (2015) study *Internet Service Provider* Networks using secondary, commercial, data on Autonomous Systems, from Telegeography, to identify the major global *Internet Service Providers*.

In the economic literature, Choi, Galeotti and Goyal (2014) study, in a game-theoretical setting, the key role played by Network metrics to model market power in communication networks. However, there is a significant lack of research analysing structural bottlenecks of the upstream Internet market using *Complex Network Analysis*, as well as its application to investigate developing country's mobile broadband operator upstream connectivity. D'Ignazio and Giovannetti (2006; 2009) study market concentration using CAIDA datasets, in the upstream layers of the Internet access markets, mainly focussing on the role of the *Betweenness Centrality* Network metric, to capture the degree of unavoidability that each Autonomous System has in this upstream access markets.

The only work applying Complex Network Analysis to the analysis of the upstream Internet market structure from an Internet periphery perspective is, as far as we are aware, by Giovannetti and Sigloch (2015). Their work shows the significance of a key network metric, the *Clustering Coefficient*, in defining the degree of hierarchical organisation of this set of internetworking relationships. Given the socio-economic relevance and impact of affordable access to the Internet, this work identifies an urgent need to exploit these initial multi-disciplinary insights further, linking the *Complex Network Analysis* of upstream connectivity data with insights from both the fields of Economics and the Development Studies. To date, and to the best of our knowledge, no research has studied the structural and bottlenecks features in the upstream Internet access market in India (or the state of Tamil Nadu), using Network Analysis methods based on data collected through active Internet periphery measurements. Hence, we indicate that our case study addresses a number of relevant and urgent gaps in the literature.

By considering this under-researched niche, we embark on an exploratory journey aiming to understand the complexity of structural connectivity bottlenecks in the upstream access

layers of the Internet in a developing country and their potential effects on end-users' affordability. We do so by using crowdsourced *traceroute*-based active Internet periphery measurements, collected by the author in the southernmost Indian state of Tamil Nadu. By using and analysing these data, we aim to explore the reliance of Tamil Nadu's mobile broadband operator's connectivity on specific *Internet Service Providers*, and the ensuing bargaining power these *ISPs* exert on the mobile broadband operators. Ultimately, the purpose of these research efforts shall result in practical policy recommendations that help the government of Tamil Nadu, and possibly India and other developing countries. This shall bridge existing levels of Digital Divide and the associated lack of mobile Internet access resulting from present high levels of this district's income disparities. The efforts of this dissertation are motivated by the aim of improving the likelihood of a future scenario, where every world citizen, independently of their gender, income or origin, has equal, affordable and educated mobile broadband access to the most valuable resources on the Internet. This access is key for providing freedom to seize new opportunities for socio-economic growth, well-being and ultimately self-fulfilment. This scenario may appear utopian, when considering the recent developments in the United States, where the Federal Communication Commission (FCC) tilted the Open Internet Order (OIO) that gives *Internet Service Providers* more control over the access and content distribution on the Internet (Forbes, 2017). By the end of 2017, the FCC then approved measures to remove the net neutrality rules (The Verge, 2017). Given the ever-increasing power gap between the small and larger International *Internet Service Providers*, we advocate the need for an increased policymakers' awareness in considering the relevance of structural bottlenecks and hierarchical ordering in the upstream Internet market and their implications on providers' connectivity and end-users' affordability. However, it is beyond the scope of this dissertation to investigate the vastness of the literature in the three fields considered, of Economics, Computer Sciences and Development Studies. Our work seeks to find meaningful associations to harness the creation of an emerging, yet crucial, field of research. Furthermore, and given the pragmatic nature of our case study, we do not aim to seek the truth or generalisability of findings, but we limit ourselves to the indication of these findings' practical implications. Our exploratory limitations are in line with this approach, suggesting that further research would provide additional explanatory power.

After this introduction, Chapter 2 aims to provide a more in-depth understanding of the research problem. More precisely, it covers relevant and insightful work in the disciplines

of Development Studies, Internet Economics and Network Science. Our aim is to identify valuable gaps by combining these disciplines' literature, which ultimately informs our contribution to knowledge. Hence, the results of Chapter 2 provide meaningful abductive inferences (stated as *Working Hypotheses*) based on observations and insights from the Literature Review. Next, Chapter 3 sets out the methodological assumptions that define the nature, scope and limitations of this dissertation. Moreover, this chapter will include a thorough description of the data collection as well as the chosen research methods, building a basis for applying the powerful concepts and routines of *Complex* and *Statistical Network Analysis* in the following chapters. Chapter 4 aims to provide a general understanding of the collected data, through the exploratory analysis of the structural features of the Tamil Nadu mobile broadband operator networks, using *Complex Network* and *Graph Visualisation Analysis*. Chapter 5, then, aims to test the *Working Hypotheses* inferred in Chapter 2 and to extend the initial descriptive results obtained in Chapter 4, by building a set of econometric models, to derive the relevant parameters' inferences. Hence, Chapter 5 performs a full *Statistical Network Analysis* of the upstream connectivity network of Tamil Nadu mobile broadband operators. This shall ultimately provide greater confidence to form pragmatic judgements about our *Working Hypotheses* abducted Chapter 2. Lastly, Chapter 6 summarises the main findings, placing our contributions to knowledge into the existing research context. This will allow us to conclude in Chapter 7 with the most valuable implications that can be derived for policy and practice, also in view of helping the Indian regulatory and telecommunication authorities when called to assess their possible actions to bridge the present digital divide.

2 LITERATURE REVIEW

‘When goods are digital, they can be replicated with perfect quality at nearly zero cost, and they can be delivered almost instantaneously. Welcome to the economics of abundance.’ (Brynjolfsson, 2013).

Instantaneous delivery of digital goods is a challenging endeavour for mobile broadband operators in developing and emerging countries. End-users simultaneously access the Internet and their connections traverse small and large *Internet Service Providers* to reach the desired connectivity destinations. The presence of too many end-users when there is too little infrastructure bandwidth is problematic for the proper functioning of the Internet at high quality while striving for a low cost. This is especially challenging in developing and low-income countries like India where the end-user access to the Internet is greatly affected by urban-rural income and gender disparities in mobile broadband access affordability (see e.g. Selvabaskar et al., 2016), education and social attitudes towards technology. In 2016, nearly 6 billion people do not have access to high-speed Internet, ‘making them unable to fully participate in the digital economy’ (WDR, 2016, p.xiii). Fair and affordable access to the world’s digital goods via the Internet can greatly influence economic growth (GSMA, 2016), inclusion (Broadband Commission, 2016), equality and social impact (WDR, 2016). A factor that strongly influences the delivery of such digital goods through the Internet value chain is the formation and functioning of the upstream Internet market structure, representing economic relationships between *Internet Service Providers*. Interestingly, national and international Internet policies focus more on the demand side of the Internet access (affordability, safety and openness), while the supply-side and even more its infrastructural aspects are often neglected (WDR, 2016).

This Chapter reviews the relevant literature from three separate fields of research: Development Studies, Network Science and Internet Economics. Our aim is to explore the relevant overlaps emerging from these three fields, which shall lead to the abduction of valuable *Working Hypotheses* (see section 3.2). Hence, we organise this Chapter thematically, the different sections start with more high-level, or historical, concepts before gradually moving to more specific and applicable studies bearing greater relevance

for the assessment of our research problem.

In the following sections, we are going to extract key insights, derived from multidisciplinary research disciplines, given their relevance to explore the upstream Internet access market features of Tamil Nadu's mobile broadband providers. These insights ultimately lead to the abduction of the relevant *Working Hypotheses* in section 2.5.

2.1 Upstream Interconnections

2.1.1 Nascent Stages of the Internet

The Internet as a whole can be seen as a scientific and technological paradigm shift (Dosi 1982), brought out through research and innovations emerged in the early stages of the computer age. Its beginnings can be traced back to research advancements at the United States Defense Advanced Research Projects Agency (DARPA). The research at DARPA led to the development of the Advanced Research Projects Agency Network (ARPANET), which represented a testing network that aimed to link universities and research institutes in the late 1960s (Cerf and Cain, 1983). At the end of 1969, ARPANET had four early linked member institutions, the UC Los Angeles, the Stanford Research Institute, the University of Utah and the UC Santa Barbara. The advancement of the ARPANET led to the creation of a number of other packet switching networks (see Glossary) such as the MFENet from the US Department of Energy, the SPAN from NASA Space Physicists or CSNET, a network for the academic Computer Science community. CSNET heralded the start of the Internet when the US National Science Foundation (NSF) granted the expansion of the CSNET to establish further links to Supercomputing Centres and other Research Networks at no additional costs for its current members. Due to this funding, the CSNET was renamed NSFNET. In 1992, the NSFNET created an Acceptable Use Policy (AUP) to enforce that the packet switching network was only used for research and education purposes in the sciences and engineering sectors (NSFNET, 1992). During the establishment of this AUP, the first *Internet Service Providers (ISPs)* such as PSINet or CERFnet introduced the first links to establish commercial traffic. The packet routing of these commercial connections would therefore pass through the NSFNET backbone, while adhering to their Acceptable Use Policies. Throughout the years it became more and more apparent that the NSFNET model ought to be replaced with a commercially operated packet switching network,

where users would need to purchase access. This vision was enforced by the TCP/IP network protocol inventor Vint Cerf (Cerf, 1990). Finally, the NSFNET backbone (see Glossary) transitioned onto a new architecture and was decommissioned in 1995. This created the necessary freedom to carry commercial traffic, a prerequisite for the privatisation and unprecedented growth of the Internet until today.

2.1.2 Internet Service Provider Relationships

The Internet can be defined as a dynamic and self-organised network of inter-connected networks. The ecosystem of the Internet is said to involve agents with a diverse set of functional roles and objectives. Hence, the Internet is composed of thousands of *Internet Service Providers* that are operating different parts of this Information and Communication Technology (ICT) infrastructure, providing services to access and use the Internet. The services of *ISPs* are typically used by content providers and end-users (including machines or ‘bots’, which are using the largest part of the Internet infrastructure). The agents composing the Internet ecosystem include:

- Access Providers: *Internet Service Providers* selling Internet Access to individuals and / or business customers (e.g. mobile broadband operators such as Aircel or Vodafone).
- Transit Providers: considered as geographically distributed large backbone network operators, which were historically, and are still presently, paid to transfer traffic over large distances. A transit provider might also be an access provider (e.g. Level 3 Communications or Cogent Communications).
- Content Providers: *Internet Service Providers* that generate the content for end-users on the Internet. Content Providers include providers of information, video, e-Commerce, social networking or search results, amongst others (e.g. Google, Facebook or Netflix).
- Content Distribution Networks (CDN): these are *Internet Service Providers* that store customer content locally for a quicker fulfilment of download requests from nearby users (e.g. Akamai). Their customers are usually access providers.
- Internet Exchange Points (IXP): IXPs operate facilities of (paid) inter-connection, where other *Internet Service Providers* may be present and inter-connect with other *ISPs* (e.g. the London Internet Exchange (LINX), or the Amsterdam Internet Exchange (AIX)).

Internet Service Providers engage in a set of formal and / or informal relationships with each other, providing access to value added end-user services, by collectively routing Internet traffic. Such relationships (and especially their routing policies, see section 2.3) determine constraints to the respective paths through which Internet traffic might flow and therefore have implications on the robustness and further engineering of the Internet.

Insight 1: Given the coexistence of formal and informal relationships amongst *Internet Service Providers*, we expect their clear identification to be quite challenging. While routing policies are often available (see above), ISPs are likely to maintain confidentiality about their business interconnection practices. Hence, we argue that available (secondary) data still contains many unidentified *Internet Service Provider* relationships, leading to the present difficulties, for current research, in providing a satisfactory picture of the Internet infrastructure.

The carriage of traffic on the Internet is usually organised through packet switching networks. This method relates to the transfer of data, which is split into smaller data packets for simultaneous transaction purposes. The carriage of these data packets through a network follows approximately symmetric transactions. This means that the senders and receivers of a given data packet are involved in the same amount of data transactions. Hence, these transactions may involve symmetrical payments, where both the sender and the receiver pay transaction costs to their *Internet Service Providers*. As described in detail below, an *ISP* may then pay their partners for the transit of the data packet in the upstream Internet, assuming they do not have a peering relationship with them (Woodcock, 2003). Following this organisation, the data packets are routed through vertically related Autonomous Systems, belonging to one or more *Internet Service Provider(s)* along the path to the final data packet destination. Internet routing mechanisms facilitate this exchange of data packets through computer networks using the so-called TCP/IP stack, which was jointly developed by Cerf and Kahn (1974) through DARPA funding, representing a cornerstone of the information-based Internet. In detail, the Internet Protocol establishes interconnections with the aim of delivering given data packets from sender to receiver, while both these agents obtain a unique Internet Protocol (IP) address. Such IP addresses are organised in address ranges (prefix) managed by Autonomous Systems. Hawkinson and Bates (1996, p.2) suitably define an Autonomous System as

‘... a connected group of one or more IP prefixes run by one or more network operators which have a single and clearly defined routing policy’.

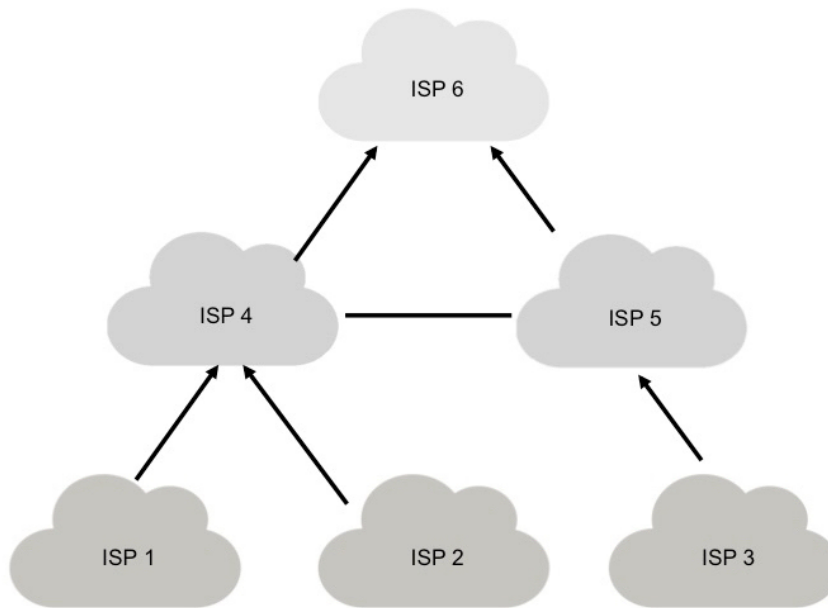
Each Autonomous System is therefore a physical collection of bare-metal (gateway) routers, represented by a unique IP address prefix (see Glossary). These IP address prefixes are under one common administrative control by the Autonomous System. However, multiple *Internet Service Providers* may share the ownership, and hence the administrative control, over an AS. Such policy-based routing techniques between IP address prefixes represent interconnections that are established between a pair of Autonomous Systems in packet switching networks, the ‘lifeline of telecommunication services’ (TRAI, 2016f). The International Telecommunications Union refers to peering or transit as relationships between Autonomous Systems (ITU, 2007, pp.7-9). Peering relates to an exchange of traffic between a defined set of Internet Protocol networks usually at no charge, except for paid peering (Norton, 2011). This exchange of traffic takes place mostly when Autonomous Systems share the same traffic volume characteristics. In any peering relationship, both sides agree to the peering conditions, which might include network coverage, operations, and maintenance of the network, as well as the volume of traffic that can be exchanged. While the process for engaging in peering relationships is often undisclosed, some *Internet Service Providers* share the peering policies for their Autonomous Systems more openly, see e.g. the peering policy of the Swiss National Research and Education Network (SWITCH, 2016). Some efforts are undertaken to collect the routing policies to the Internet through the Internet Routing Registry (IRR), a distributed routing information database formed in 1995 (IRR, 2016). The distribution and organisation of IP addresses is managed at continental level through the Réseaux IP Européens Network Coordination Center, the regional Internet registry for Europe, which also collects routing information using RCC, the Remote Route Collector (RIPE NCC, 2016a). The Asia Pacific Network Information Center (APNIC), the Regional Internet Registry of Asia, also engages in the collection of routing information (including Indian ones) through the above-mentioned Internet Routing Registry (APNIC, 2016). Nevertheless, when the ASes of two *Internet Service Providers* enter into a relationship, the type of contract arrangements depends largely on the balance of contributions that benefit both parties (ITU, 2007). If a peering arrangement is not possible, the *Internet Service Provider* parties might engage in transit arrangements, which allow them to reach all remaining parts in the Internet periphery. Here, larger *Internet Service Providers* sell access to their network to their customer AS networks of

other *ISPs*. Once a transit arrangement is set, the sender pays the full costs of interconnection. The charges for these interconnections are usually undisclosed and negotiated on commercial terms. As previously shown, transit arrangements with one of the large Tier-1 *Internet Service Providers*, or those that directly connect to the Internet backbone, can provide a smaller *Internet Service Provider* with access to the rest of the Internet, while also introducing costs leading to potentially high global connectivity prices (ITU, 2007, p.9; CAIDA, 2016a). By gaining such access to the rest of the Internet, a smaller *Internet Service Provider* would be reliant on those larger *Internet Service Providers* for the purpose of global internetworking to and from the Internet periphery, interconnecting end-users. The small number of large upstream *Internet Service Providers*, with a strong interconnection demand from downstream *ISP,s* forms the higher hierarchical structure of the Internet, the core, holding great negotiating power over interconnection practices and prices with the smaller downstream ones (D'Ignazio and Giovannetti, 2006, pp.2-13). Peering is common amongst members of the Internet core who also provide global connectivity through paid transit to the other *ISPs*, forming the lower layers (Tiers) of the Internet (Woodcock, 2003). Furthermore, the oligopolistic structure of the Internet core guarantees the largest *Internet Service Providers* with unidirectional revenue flows arising from transit payments from smaller *Internet Service Providers* residing in the Internet periphery. This asymmetric flow of resources reinforces the incentives to minimize transit costs for a growing number of *Internet Service Providers* and end-users in the Internet periphery and maximises their set of peering relations.

A large body of literature in Computer Science focuses on the study of the relationships between Autonomous Systems to explain peering and transit relationships. The relationships amongst ASes are considered to have a significant impact on the flow of traffic through the Internet (Subramanian et al., 2001). The Autonomous System roles in these relationships are, according to Alaettinoglu (1996) and Huston (1999), either of *provider-to-customer*, *customer-to-provider* or *peer-to-peer* nature. This definition is also used in the work of Gao (2001, p.734), who further states that two Autonomous Systems, which are operated by one *Internet Service Provider*, may have *sibling* relationships, where each AS provides transit services for the other. This is especially relevant when considering that one *Internet Service Provider* might operate multiple Autonomous Systems. Moreover, Gao and Rexford (2000) show that a pair of ASes may peer indirectly through a transit Autonomous System. Gao, Griffin and Rexford (2001) expand on this,

suggesting that a pair of Autonomous Systems may also have backup relationships to provide connectivity in the event of failures or downtimes.

Nevertheless, the structuring of the Internet allows larger Tier-1 providers to obtain central positions on the Internet, given their transit relationships that are crucial for interconnecting traffic, across geographical distances, with / for other *Internet Service Providers*. According to Economides (1995, p.678), in economics, structural bottlenecks in the interconnections occur when an economic agent has a monopoly, and market power, over a link (or relationship) with other economic agents, creating essential facilities within a network. Structural bottlenecks, hence, cause traffic flow congestions in the digital supply chain, which occur when an Autonomous System receives more data traffic than it can cope with. This definition provides a cornerstone for our work. When referring to ‘structural bottlenecks’, we speak of the economical, rather than the technological, definition. This view of Economides (1995) is also identified by Subramanian et al. (2001), stating that relationships between Autonomous Systems have a significant impact on the flow of traffic through the Internet, while hierarchy symbolises business relationships between Autonomous Systems. Subramanian et al. (2001) also state that customers should be at a lower hierarchical layer compared to their providers, a concept which is best captured by using methods of directed network graphs where edge directions indicate the types of relationships between two Autonomous Systems. To better capture the types of relationships between ASes, Luckie et al. (2013) propose the usage of the concept of an Autonomous System’s *Customer Cone* based on the PhD work of Giotsas (2014). The following Figure 2-1 visualises this concept of a *Customer Cone*.



Key

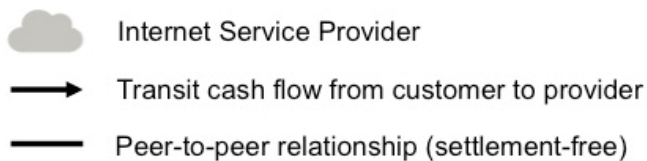


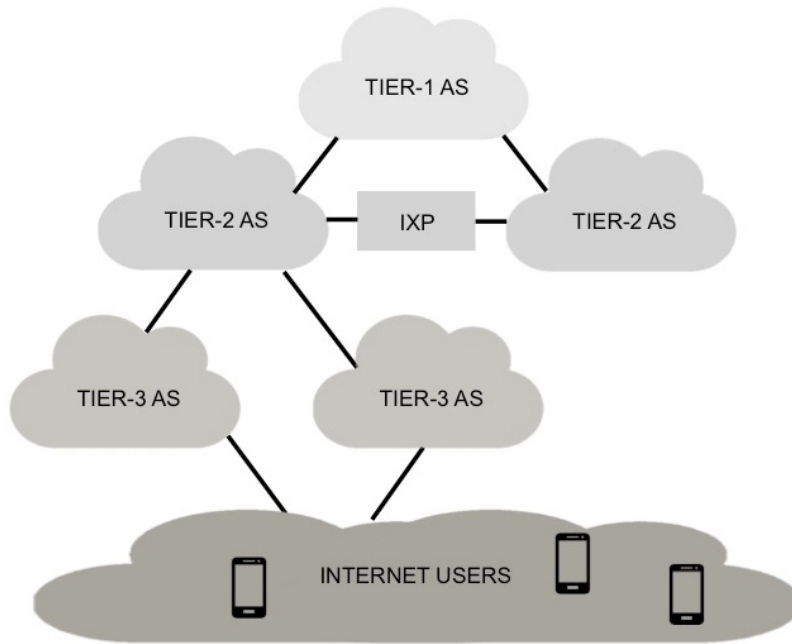
Figure 2-1: Internet Service Provider (ISP) Customer Cones.

Considering Figure 2-1 above, the customer cone of ‘ISP4’ would be ‘ISP1’, ‘ISP2’, whereas the customer for ‘ISP5’ would be ‘ISP3’. This is especially relevant since two Autonomous Systems might have a *peer-to-peer* relationship in one location of the Internet and a *provider-to-customer*, or *customer-to-provider* one at another location on the Internet. Nevertheless, and despite this complexity, the business relationships amongst pairs of ASes is a theme where too little economic research has been undertaken to date. However, we consider these relationships as hugely important when studying the formation and structure of the upstream Internet market of mobile broadband operators, since Internet connectivity is not established on the basis of the shortest paths to the final destination but on their economic value.

The fact that an *Internet Service Provider* may inhabit one or a multitude of Autonomous Systems, whereas an AS may also be shared amongst a set of *Internet Service Providers*,

adds complexity to the interconnection relationships. This is further complicated by the fact that there is no central authority managing the total number of connecting Autonomous Systems. The organisation of these Autonomous Systems is only vaguely defined. *Internet Service Providers* that manage ASes usually also provide global connectivity to their customer networks, but this type of connectivity comes in a variety of sizes and structures (MIT, 2009). The shared definition amongst Computer Science practitioners classifies *ISPs* into three different Tiers, as Figure 2-2 below illustrates.

Here, Tier-3 *Internet Service Providers* are believed to provide connectivity to a low number of geographically local end customers while being connected to upstream Tier-2 *ISPs*. These Tier 2 *Internet Service Providers* cover the regional connectivity scope (state, or region wide), while linking to Tier-1 *ISPs* for international connections. Hence, Tier-1 *Internet Service Providers* capture a global connectivity scope, being able to reach any Autonomous System on the Internet (MIT, 2009). Nevertheless, there are only a handful of these large International Tier-1 *Internet Service Providers*, stated e.g. in the CAIDA (2016a) AS-Rank. However, the specific literature finds no consensus about the amount of such large Tier-1 *ISPs*. Importantly, Tier-1 *Internet Service Providers* are not reliant on buying connectivity transit services from other *ISPs* but mostly rely on settlement-free peering relationships with other large *Internet Service Providers*, reciprocally exchanging traffic between each other. This provides them with bargaining power and global connectivity criticality on the Internet. Moreover, the transit services of large Tier-1 *ISPs* would usually cover priced services that allow smaller *ISPs* (from Tier-2 or Tier-3) to access the entire Internet through routing agreements. These routing agreements may not be transitive since *Internet Services Providers* are not obliged to carry traffic to other *ISPs*. Moreover, Tier-2 *Internet Services Providers* are usually peering with some other *ISPs* but are still reliant on purchasing Internet Protocol (IP) transit (see Glossary) from Tier-1, or other regional Tier-2 *ISPs*, depending on the final data packet destination. An *ISP* that purchases transit would then be a customer in a *customer-to-provider* relationship, as described in their routing policy above. Such a Tier-2 *ISP* would most likely try to save IP transit costs by establishing peering relationships with as many Tier-2 or Tier-1 *Internet Service Providers* as possible. Lastly, Tier 3 *ISPs* do not usually sell any transit to other *ISPs* but are entirely reliant on purchasing transit from other *Internet Service Providers* in order to reach the entire Internet. The following Figure 2-2 provides an overview of possible interactions between *ISPs* and an Internet Exchange Point (IXP) located within the three Tiers.



Key

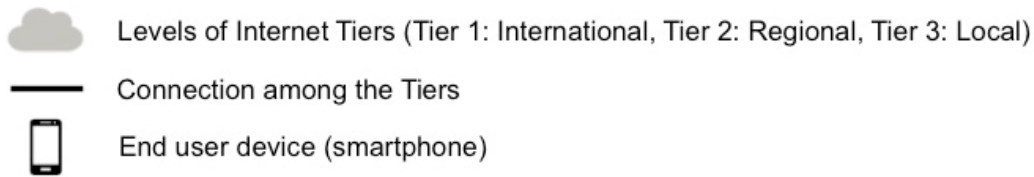


Figure 2-2: Internet Protocol (IP) Tier networks.

Insight 2: Based on the economic nature of *Internet Service Provider* relationships, we expect to observe a hierarchical network structure where a low number of globally acting Tier-1 *Internet Service Providers* provide global connectivity to a larger number of regional Tier-2 and local Tier-3 *Internet Service Providers*, except for when a given data traffic remains local. While this economic nature is largely agreed upon in the literature (e.g. Luckie et al., 2013), little research seems to focus on the implication of such hierarchical structuring, characterised by strong bargaining powers of a few Tier-1 *Internet Service Providers*. Our work aims to provide insights into this direction.

2.2 Mobile Broadband and Digital Divide in India

2.2.1 Indian Internet Infrastructure

Similar to other countries, the rise of the Internet infrastructure in India started with the launch of their Educational Research Network (ERNET) as a joint effort of the Indian Department of Electronics and the United Nations Development Program, in 1986 (ERNET, 2016). The network of ERNET interconnected eight institutions including the Indian Department of Electronics, the five Indian Institutes of Technology (IIT) at Delhi, Bombay, Kanpur, Kharagpur and Madras (Chennai), the National Centre for Software Technology in Bombay and the Indian Institute of Science in Bangalore. Later in 1995, India joined the commercial Internet when Videsh Sanchar Nigam Limited (VSNL) formally launched their Gateway Internet Access Service (GIAS) in Bombay, Delhi, Kolkata and Chennai with the following message:

*‘VSNL India’s Gateway to the world welcomes you to surf the cyberspace’
(VSNL, 1996, p.23).*

The access to the VSNL gateway came at a cost of INR 25,000 (approx. US\$ 389.18 (XE.com, 2016)) for 250 hours of TCP/IP accounts at a speed of 9.6 kbps (kilobits per second). In just six months, VSNL added 10,000 Internet users to the Gateway Internet Access Service, while access was limited to New Delhi, Mumbai, Kolkata and Chennai (Fennell et. al., 2016). As of July 2016, India is estimated to have reached 46.2 crore (462 million) Internet users, representing a penetration level of roughly 34.80% of the Indian population, according to the Internet and Mobile Association of India (IAMAI) report stated in Internet World Stats (2016). The penetration rate mentioned by Internet World Stats (2016) corresponds to India having 25.80% of all Asian Internet users. However, the available statistics on the Indian Internet access and penetration rates vary wildly in the academic, business and governmental sources. As of January 2016, Statista (2016a) reports that India has a considerably lower number of 375 million Internet users, compared to the 462 million in the Internet World Stats (2016). For 2015, the International Telecommunications Union country profile for India marks that 26.00% of the individuals are using the Internet (ITU, 2015d). This corresponds to approximately 340.86 million Internet users, when combined with the World Bank (2016a) Indian population statistics during the same year. The World Bank (2016a) itself states that India had 43.99 Internet users per 100 inhabitants in 2015, which would correspond to

staggering 576.81 million Internet users. Overall, the statistics on the number of Internet users shows large variations depending on the source. Nevertheless, these differences are not surprising considering that a larger number of the Indian population accesses the Internet through shared access or Cyber Cafés, representing a dominant share of 37.00% of Internet access back in 2010 (TRAI, 2010, p.24). Another method for accessing the Internet is through Kiosk operators in rural panyachats, providing wireless connections through accessing the (incumbent) telecom provider networks from county towns (Jhunjhunwala, Ramachandran and Bandyopadhyay, 2004). These examples show that Internet access solutions for rural India are especially creative, which makes it very challenging to trace the number of end-users accessing the Internet. One solution to this measurement issue is the recently introduced ‘Aadhar’ digital identifier by the Indian government. Since every mobile broadband operator is obliged to make use of this identifier, we expect Internet access statistics to become more transparent in the upcoming years.

Globally, approximately two-thirds of the 3.2 billion people that were online by the end of 2015 were from developing countries as estimates of the International Telecommunications Union state (ITU, 2015a). However, 4 billion people (two-third of the world’s population) remained offline in 2015 (ITU, 2015a). Moreover, most of the offline population resides in developing countries such as India. However, International Organisations have not yet reached a full consensus on whether India should still be considered as a developing country: while the United Nations (2016) considers India as a ‘developing nation’ in their Standard Country and Area Codes for Statistical Use, the World Bank (2016b) raised India’s ranking to a ‘lower-middle income country’ in 2015. Considering India’s number of Internet users compared to their total population of 1.311 billion in 2015 reveals that also a large portion of the Indian population remains offline (World Bank, 2016a). The Indian population that could access the Internet were mostly using wireless, rather than wireline, means of narrowband and broadband connections, where wireline corresponds to wired broadband technologies and wireless to cordless ones. The Telecom Regulatory Authority of India (TRAI, 2016c) statistics provides reliable data on the means of Internet connectivity usage since each subscription corresponds to a registration. TRAI (2016c) notes in their Performance Indicator report that India reached a total number of 345.60 million Internet subscribers in 2016, representing a different statistic to the number of Internet users. However, the key variation in these statistics seems to be based on the differences between ‘broadband

subscription’ and ‘broadband user’, since a multitude of users is frequently sharing a single subscription with their families and peers. The actual number of Indian Internet users remains a mystery until the ‘Aadhar’ identifiers system is well established. Hence, we will use the Internet usage data given by the Telecom Regulatory Authority of India as illustrated in Table 2-1 below. In detail, the total number of Indian Internet subscribers corresponds to 21.26 million wired Internet subscribers, 0.62 million Fixed wireless Internet subscribers (Wi-Fi, Wi-Max, Radio or VSAT connections) and staggering 345.60 million mobile wireless Internet subscribers in September 2016 (TRAI, 2016c, p.28). Moreover, the number of total Internet subscribers is divided between 247.69 million urban subscribers and 119.79 million rural ones. On a per 100 inhabitant basis, every second urban-living Indian (61.98%) have a smartphone subscription but for rural areas, this figure declines, where circa one in ten (13.65%) people has a smartphone subscription (TRAI, 2016c, ii), representing 44.19 million rural Internet subscribers.

Indian Internet subscriber base per segments in September 2016	
Internet subscriber segments	Number of Subscribers in millions (TRAI, 2016c)
Total Internet subscribers	367.48
Wireless Internet subscribers	346.22
Mobile Wireless (Phone + Dongle)	345.60
Fixed Wireless (Wi-Fi, Wi-Max, Radio & VSAT)	0.62
Wired Internet subscribers	21.26
Broadband subscribers	192.30
Narrowband subscribers	175.18
Urban Internet subscribers	247.69
Rural Internet subscribers	119.79
Urban Broadband	148.11
Urban Narrowband	99.58
Rural Broadband	44.19
Rural Narrowband	75.60

Table 2-1: Indian broadband subscriber base per segment, Source: TRAI (2016c).

2.2.2 Mobile Broadband in Tamil Nadu

Mobile broadband generally refers to wireless Internet access through portable devices such as smartphones or dongles (TRAI, 2016c) under a given speed restriction (see Glossary). Mobile broadband operators such as Aircel or Vodafone are access providers that offer mobile broadband services at price plans that usually vary regarding data usage allowances (e.g. 200MB), connectivity speeds (such as 2G, 3G, 4G) and added services (e.g. unlimited Facebook access). Some of these service variations such as Quality of Service (QoS) indicators or connectivity speeds remain mostly invisible to an end-user. ITU (2003, p.9) early defines mobile broadband Internet access through mobile broadband operators as:

‘Internet connections that are significantly faster than today’s dial-up technologies, but it is not a specific speed or service.’

This early definition deliberately excludes connection speeds or added services. However, Policy practitioners from a range of countries such as the United States of America, Brazil, and Bangladesh often include ‘to-be-delivered’ connection speeds in their mobile broadband policy definitions (TRAI, 2016a, pp.4-5). A practical disadvantage of these definitions (with connection speeds) is that policies ought to be constantly revised in order to stay up-to-date with the pace of technological advancements. This is also the case for the definition of the Department of Telecommunications at the Government of India who defines ‘Broadband’ (effective from January 2015) in TRAI (2016a, p.2) as:

‘a data connection that is able to support interactive services including Internet access and has the capability of the minimum download speed of 512 kbps to an individual subscriber from the point of presence (POP), of the service provider intending to provide Broadband service’

while demanding an increase of the broadband speed in the definition to 2 Mbps (TRAI, 2016a, p.5). Broadband speeds are usually divided into upload and download speeds, while download speeds are roughly double the defined upload speeds (TRAI, 2016a). However, the broadband definition of the Department of Telecommunications at the Government of India seems considerably different to those adopted by researchers in the field of Information and Communication Technology for Development (ICT4D). In their ‘Building Broadband’ report, Kim, Kelly and Raja (2010) of the Global Information and Communication Technologies Department at the World Bank argue against mobile broadband definitions that cover network connectivity and minimum transmission speeds. Moreover, they propose a mobile broadband definition as an ecosystem, which would involve the mobile broadband networks, the services being carried through the networks, the applications delivered, and the end-users served. Hence, Kim, Kelly and Raja (2010, pp.iv-22) consider the supply and demand sides of the Internet market as well as the access to these networks and their services as forming a unique ecosystem. This perspective seems particularly suitable for our work in this dissertation as it enhances more traditional views on mobile broadband.

In 1994, the Indian Department of Telecommunications introduced the formal organisation of Telecom Service Areas through their National Telecom Policy (GOV-IN, 2016b). As a result of this policy, India is divided into 19 Telecom Service Areas and 4 Metro Service Areas. These Metro Service Areas represent the cities of Delhi, Mumbai,

Kolkata and Chennai. Each Telecom Service Area only allows for a maximum number of access providers. This means that any Indian mobile broadband operator needs to acquire fixed-term licenses for providing their services in any of these Telecom Service Areas. The Telecom Service Area with the highest number of broadband Internet subscribers (wireline and wireless in millions) was Maharashtra with 30.62, followed by Tamil Nadu (incl. Chennai) with 27.46 and Andhra Pradesh with 27.46. The highest number of urban broadband Internet subscribers (wireline and wireless in millions) in September 2016 was Delhi with 22.27, followed by Tamil Nadu with 22.21, and Maharashtra with 20.33 (TRAI, 2016c). The Telecom Service Area with the most rural broadband Internet subscribers is Uttar Pradesh with 12.32, followed by Maharashtra with 10.30, and Andhra Pradesh with 9.68 (TRAI, 2016c). Surprisingly, Tamil Nadu is more far off with 6.97 rural Internet subscribers. India's top three service areas with respect to broadband subscriptions are Tamil Nadu (including Chennai) with 19.32 million, followed by Maharashtra with 18.13 million and Andhra Pradesh with 16.90 million.

Tamil Nadu Mobile Broadband Infrastructure

To explore this in more detail, the Internet subscriber base (wireless and wireline in million) in the state of Tamil Nadu (including the urban area of Chennai) reached 29.18 (19.32 broadband, 9.86 narrowband) by September 2016. Based on a survey of 45,435 respondents from 40 countries in May 2015, Pew Research (2016) reports that only 17.00% of Indians own a smartphone. Moreover, 27.00% of the age group 18-34 reported owning a smartphone, and only 9.00% of respondents aged over 35 years. Pew Research (2016) also finds a correlation between the number of smartphone users and education as well as income.

The evidence and data discussed in the following paragraphs are obtained from the Indian Telecom Services Performance Indicator Report (July – September 2016) of TRAI (2016c). According to TRAI (2016c), the Telecom Service Area of Tamil Nadu currently covers four mobile broadband operators namely Aircel, Bharti Airtel, BSNL and Vodafone. Icompare (2016a; 2016b) states that the same mobile broadband operators provide their mobile broadband services also in the Chennai Metro Service Area. These Internet subscribers are comprised of 22.21 million urban subscribers and 6.97 rural subscribers in September 2016 (TRAI, 2016c, p.30), where 13.61 million represent urban broadband subscribers, compared to 3.67 million rural ones. The narrowband subscribers correspond to 8.59 million urban and 3.31 rural ones. This relates to 51.60% of the urban

population having access to the Internet and only 24.68% of the urban one (TRAI, 2016c, p.35).

Table 2-2 below depicts the total Indian broadband (Wireline and Wireless) subscriber base for each of the mobile broadband operators with a Tamil Nadu presence, providing an idea of their total network size and connectivity differences.

Indian total subscriber base (wireline and wireless) 2016 for Tamil Nadu mobile broadband operators		
Mobile Broadband Operator (with Tamil Nadu Presence),	Subscriber Base QE June 2016 (total India in millions)	Subscriber Base QE September 2016 (total India in millions)
Aircel	88.93	90.14
Bharti Airtel	255.73	259.94
BSNL	89.54	93.77
Vodafone	199.38	200.72
Key QE: End of the quarter		

Table 2-2: Indian total subscriber base (wireline and wireless) 2016 for Tamil Nadu mobile broadband operators, Source: TRAI (2016c).

All of the four Tamil Nadu mobile broadband operators except BSNL provide services in all Indian states, covering GSM and CDMA services. BSNL has no broadband service operating presence in the metro service areas of Delhi and Mumbai. The wireless subscriber base in Table 2-3 below shows only the wireless broadband subscriptions, compared to the previous Table 2-2 above, which includes the wireline broadband subscriptions. The differences are marginal, showing the importance of mobile broadband subscriptions for each of the four operators. TRAI (2016c) notes that 94.05% of all Indian Internet subscriptions are mobile wireless ones (5.78 wired and 0.17 fixed wireless ones). This is also the case for Tamil Nadu. Fennell et al. (2016) find, based on a sample of 38 survey respondents across rural districts of Tamil Nadu (Vellore, Madurai and Pudukkotai), that 68% of all respondents access the Internet through Mobile devices (17% not using, 10% Laptop, 5% Computer).

Indian total wireless subscriber base 2016 for Tamil Nadu mobile broadband operators		
Tamil Nadu Broadband Operator	Wireless Subscriber Base QE June 2016 (total India in millions)	Wireless Subscriber Base QE September 2016 (total India in millions)
Aircel	88.93	90.14
Bharti Airtel	255.73	259.94
BSNL	89.54	93.77
Vodafone	199.38	200.72
Key QE: End of the quarter		

Table 2-3: Indian total wireless subscriber base 2016 for Tamil Nadu mobile broadband operators, Source: TRAI (2016c).

Table 2-4 below provides an overview of the rural market share (wireless subscriber base, GSM and CDMA) for each of the broadband operators with Tamil Nadu presence. Given the TRAI (2016c) data, Bharti Airtel is the rural market leader but Vodafone accounts for more than half of their subscribers from Indian rural areas, showing some of the strategic differences amongst the mobile broadband operators.

Indian rural market share in 2016 for Tamil Nadu mobile broadband operators		
Tamil Nadu Broadband Operator	Rural Market Share (wireless subscribers) in % end of September 2016	Rural subscribers in % of the total mobile broadband subscribers
Aircel	6.98	34.54
Bharti Airtel	28.40	48.72
BSNL	6.93	32.94
Vodafone	23.86	53.01
Key GSM: Global System for Mobile Communications		

Table 2-4: Indian rural market share in 2016 for Tamil Nadu mobile broadband operators, Source: TRAI (2016c).

The total number of the GSM wireless subscriber base of the four mobile broadband operators is reported in Table 2-5 below. Here, Aircel shows the largest subscriber base, followed by Bharti Airtel, Vodafone and BSNL.

Indian GSM wireless subscriber base in 2016 for Tamil Nadu mobile broadband operators		
Market No.	Tamil Nadu Broadband Operator	GSM wireless subscriber base (September 2016)
1	Aircel	21,788,965
2	Bharti Airtel	18,028,189
3	Vodafone	15,874,774
4	BSNL	8,802,836

Table 2-5: Tamil Nadu market share 2016 for Tamil Nadu mobile broadband operators, Source: TRAI (2016c).

TRAI (2016c) still lacks 3G subscription data, which would have been valuable here. Nevertheless, we expect the number of 3G subscriptions to be much lower, potentially due to affordability and access reasons.

Insight 3: Given the statistics above (e.g. TRAI, 2016c), Tamil Nadu (incl. Chennai) represents an Indian state with a high number of urban broadband subscribers, but a low number of rural ones. This makes Tamil Nadu very interesting for our case study as it allows us to examine the apparent mobile broadband affordability disparities between urban and more rural districts.

Affordability of Mobile Broadband

Generally, the World Economic Forum considers affordable connectivity as a key infrastructural element for a robust digital economy (WEF 2017, p.11). Interestingly, Internet end-users located in the Asia-Pacific region pay higher charges for International Internet connections than their global counterparts (Sultana, 2016). A key reason for this is, that developing countries that aim to connect to the core of the Internet (where global connectivity takes place) pay the full costs of international leased lines to the country, as Sultana (2016, p.26) pointed out at an ITU workshop in New Delhi. Such costs to access the upstream Internet market must then be passed on by mobile broadband operators to their end-users, hence raising the barriers to adoption of affordable mobile broadband. Hence, especially developing and emerging economies, where most of the world's offline population resides (ITU, 2016), face these affordability challenges.

Up to date price plans for mobile broadband operators are generally not easy to obtain from a centralised source. We found no single source collection for up-to-date subscription information. One of the most comprehensive collections of Tamil Nadu price plans during our data collection period was GSMOutlook, an online resource. Here, Aircel showed 24 different price plans (GSMOutlook, 2015a), followed by Bharti Airtel with 15 (GSMOutlook, 2015b), BSNL with 13 (GSMOutlook, 2015c), and Vodafone with 11 (GSMOutlook, 2015d), see Appendices. Hence, compared to the other access providers, Aircel provided the largest choice to their Tamil Nadu existing and potential customers at the beginning of 2015 (see section 5.4.1).

Initial observations might suggest that these operator price plans seem affordable; however, a comparison of this pricing to the state income level reveals problems in relation to their real affordability. In 2014, Tamil Nadu accounted to 5.4 per cent of the total rural workers in India, while the share of rural workers to total workers in Tamil Nadu was at 50.70% (TN-GOV-IN, 2014). According to IBEF (2016), the gross state domestic product (GSDP) of Tamil Nadu grew between 2004-2005 and 2015-2016 at a compound annual growth rate of 12.31%, reaching US\$ 140.03 billion in 2015-2016. This

is equal to a per capita gross state domestic product (GSDP p.c.) of US\$ 1,941.14 at current prices. The State of Economy Chapter I in TN-GOV-IN (2014) finds that the Tamil Nadu Net State Domestic Product (NSDP) per capita income increased from INR 58,360 (approx. US\$ 908.27 (XE, 2016)) in 2012-2013 to INR 62,361 (approx. US\$ 970.54(XE, 2016)) in 2013-2014. These incomes are far higher than the Indian average per capita incomes as the following Table 2-6 below illustrates (TN-GOV-IN, 2015, p.20).

Per capita income at constant prices Tamil Nadu vs. All India				
Years	Per Capita Income in INR Tamil Nadu at constant prices	Growth rate in per cent	Per Capita Income in INR All India at constant prices	Growth rate in per cent
2004-2005	30,062	-	24,143	-
2005-2006	34,126	13.52	26,015	7.75
2006-2007	39,166	14.77	28,067	7.80
2007-2008	41,314	5.48	30,332	8.07
2008-2009	43,193	4.55	31,754	4.69
2009-2010	47,394	9.73	33,901	6.76
2010-2011	53,507	12.90	36,342	7.20
2011-2012	57,093	6.70	38,037	4.66
2012-2013*	58,360	2.22	39,168	2.97
2013-2014*	62,361	6.86	39,961	2.84
Key				
* Based on quick (2012-2013) and advanced (2013-2014) estimates				

Table 2-6: Per capita income at constant prices Tamil Nadu vs All India, Source: TN-GOV-IN (2015).

Nevertheless, the state of Tamil Nadu shows considerable gaps in per capita income depending on its different districts. TN-GOV-IN (2014) states that the district of Ariyalur,

for example, shows the lowest per capita income of INR 16,559 (approx. US\$ 257.71 (XE, 2016)). Such differences easily cause disparity problems between the urban and rural districts in the state (TN-GOV-IN, 2014). This wide inter-district disparity in per capita income represents a major policy concern that needs to be addressed (TN-GOV-IN, 2014, p.2). CNN Money (2016) states that affordability of Internet-connected devices can be challenging in India since roughly 75.00% of the population earns less than INR 5,000 (approx. US\$ 77.82 (XE, 2016)) per month. While the state of Tamil Nadu shows a higher per capita income than India in total, we argue that some districts of the state itself, as well as the city of Chennai, still face mobile broadband affordability issues. While mobile broadband pricing starts at INR 8 for 25MB per one day, monthly plans quickly charge the end-user more than INR 110 a month, or INR 1,320 per year, representing a large portion of the per capita income for the less-favoured population in rural districts in Tamil Nadu. Based on a recent survey of 340 respondents in multiple Tamil Nadu districts, Selvabaskar et al. (2016) find great monthly family income disparities (amongst an age group of 31-40 years) between the urban and rural districts of Tamil Nadu. In detail, their primary data finds a monthly family income of INR 20,001-30,000 in urban and INR 10,001 – 20,000 in rural areas (Selvabaskar et al., 2016, p.4-5). Although, their survey selection of respondents might have been biased, given the high rural income compared to TN-GOV-IN (2014). Some researchers such as Gehring and Kishore (2008) note, based on measurements of India's Gini coefficient, that income inequalities in India are fairly small, compared to other countries. Nevertheless, there seem to be great income inequalities amongst the urban and rural districts of Tamil Nadu, as Sundar (2015), based on data by the Indian Department of Economics and Statistics at the Government of Tamil Nadu, points out. This dataset represents one of the most comprehensive collections of per capita income data for Tamil Nadu. Despite dating back several years, the data were only published recently. Hence, we see a lack of up-to-date information but notice the great income disparities as also shown by the work of Selvabaskar et al. (2016). Sundar (2015) shows great annual per capita income disparities that may be associable to the different districts' economic sectors (Agriculture, Industry and Services). The lowest per capita income for the years 2010-2011 was shown by the districts of Ariyalur (INR 16,559, approx. US\$ 258 (XE, 2016)) and Perambalur (INR 17,922, approx. US\$ 280 (XE, 2016)), whereas the Tamil Nadu districts with the highest per capita income were Tiruppur (INR 72,479, approx. US\$ 1,131 (XE, 2016)) and Thiruvallur (INR 70,778, approx. US\$ 1,104 (XE, 2016)), (Sundar, 2015). This, again,

shows the large affordability disparities between Tamil Nadu's urban and rural districts.

Insight 4: The affordability of mobile broadband price plans, together with Quality of Service, seems to be very low in Tamil Nadu, potentially leading to unequal end-user access to the upstream Internet market. Since mobile broadband operator costs to access the global Internet infrastructure are likely to be passed on to end-users (Sultana, 2016), we infer that affordability of mobile broadband may be linked to the interconnection structure among the providers in the upstream Internet market (see Insight 2 above and our Working Hypotheses in section 2.5 below). Here, large International Internet Service Providers may have built positions of strong market power that influence the economic terms of connectivity for their smaller national, or regional, downstream customers.

Quality of Service (QoS)

According to an ITU (2013b) International Telecommunications Regional Group meeting in Africa, Quality of Service (QoS) measurements for broadband Internet are usually defined at a National level. Hence, we define Quality of Service according to the TRAI (2016e) definition of their Quality of Service (QoS) of Broadband Service Regulations 2014 (Second Amendment). These regulations state that the Telecom Regulatory Authority of India regularly assesses the compliance status of the mobile broadband operators for each of India's Telecom (or Metro) Service Areas, through benchmark parameters. Each of these operator's provided parameters is assessed against a benchmark metric set by the Telecom Regulatory Authority of India (TRAI, 2016b). Table 2-7 below provides an overview of Tamil Nadu's mobile broadband operators' evaluations for 2015. Interestingly, the Telecom Regulatory Authority of India might audit the mobile broadband operators, while the operators provide the necessary metrics themselves (TRAI, 2014). This creates a potential incentive for misrepresenting Quality of Service data, introducing an element of scepticism about the reliability of self-reported data.

Tamil Nadu operators Quality of Service (QoS) benchmark					
Benchmark Parameter	Technology	Aircel	Airtel	BSNL	Vodafone
Service Provisioning, within 4hrs with 95% success rate	All (2G and 3G)	99.41	92	100	99.69
Successful data transmission download attempt, > 80%	2G (GSM)	100	100	98.57	100
	3G	100	100	99.11	100
Successful data transmission upload attempt, > 75%	2G (GSM)	100	100	95.10	100
	3G	100	100	98.82	100
Minimum download speed (Kbps), to be measured for each plan	2G Plan 1	150.37	178.38	52.66	120.90
	3G Plan 1	2249.39	1357.03	1377.71	2069.41
Average Throughput for Packet data (Kbps), > 75% of the subscribed speed	2G Plan 1	198.77	212.98	106.11	164.01
	3G Plan 1	3925.33	2769.43	2244.28	6657.33
Latency, Data <250ms	2G (GSM)	77.53	228	1021.62	132
	3G	40.89	175	115.50	32
PDP Context Activation Success Rate, > 95%	2G (GSM)	98.05	99.95	99.68	99.89
	3G	98.05	100	99.83	99.61
Drop rate, <= 5%	2G (GSM)	0.98	0.68	0.41	3.66
	3G	0.82	0.18	0.28	0.49
Key hrs: Hours kbps: Kilobit per second ms: Milliseconds PDP: Packet Data Protocol					

Table 2-7: Tamil Nadu mobile broadband operator Quality of Service benchmark, Source: TRAI (2016b).

2.2.3 The Next Big Frontier

India reached staggering 1.127 billion wireless subscribers (1.112 billion GSM

subscribers) at the time of completing this dissertation (TRAI, 2017). Anticipating the growth of the Indian Mobile Broadband Sector, Ericsson (2016) predicts that India will face 1.370 billion mobile subscriptions by the year 2021. Moreover, they anticipate 810.0 million smartphones with mobile broadband subscriptions spread across the Indian population, resulting in a seventeen-fold growth of smartphone traffic until 2021. In contrast to Ericsson (2016), Statista (2016b) predicts a lower number of 317.1 million Indian smartphone users by the end of 2019. Nevertheless, both of the anticipated growth rates align with a statement by the Telecom Regulatory Authority of India. Here, TRAI (2016a) notes that most of the Indian Internet users consume data such as video content mostly via smartphones, while India's strong mobile broadband usage is highlighted in their underdeveloped wireline infrastructures. Considering the impact of India's billion mobile subscribers utilising mobile broadband services, it becomes clear that India's digital leap is just about to start. The International Telecommunications Union together with UNESCO argue in their State of Broadband Report that India is the 'Next Big Frontier' of the digital world (Broadband Commission, 2016, p.19). To reach this 'Frontier', the Ministry of Electronics & Information Technology at the Government of India launched the 'Digital India' programme described by GOV-IN (2016a), which is centred on three visions, where

- (i) The Digital infrastructure is a core utility to every citizen,
- (ii) The Governance and Services are available on demand and
- (iii) Citizens are digitally empowered.

Ericsson responded to the 'Digital India' programme through an Economic Times (2014) article stating that mobile broadband is the only platform that can deliver the 'Digital India' vision, whereas:

'The broadband infrastructure in the country needs to be expanded to offer superior coverage, quality and capacity'.

Moreover, the expansion of the broadband infrastructure includes the provisioning of new spectrum (Economic Times, 2014). Ericsson also states that:

'India must think long term in terms of laying out its National broadband policy so that supply-side constraints are managed in such a manner that 600 Million broadband subscribers can be serviced by the Year 2020.'

The coverage, quality and capacity of the mobile broadband infrastructure refer to the growing number of mobile broadband subscriptions as well as end-user's content usage patterns. We consider that the management of supply-side constraints should partially relate to affordable access for *Internet Service Providers* to the upstream Internet market, providing mobile broadband operators with the chance to offer more affordable services to their customers, required for driving the Internet's long-term socio-economic impact. The recent World Development Report states that 6 billion people do not have access to high-speed Internet, as delivering universal digital access follows investments in infrastructure and competition reforms to telecommunication markets (WDR, 2017).

2.2.4 Socio-Economic Impact of Mobile Broadband

Given its rapid development, mobile broadband and the Internet are becoming integral parts of our economies and their structural changes (OECD, 2008). The Broadband Commission for Sustainable Development states that mobile broadband is key for many development projects and considered an essential part for the delivery of the 17 Sustainable Development Goals of the United Nations Development Program (Broadband Commission, 2016) a position shared by The Internet Society (2016). The broad coverage of basic mobile broadband, especially in low-income countries, allows for driving mobile-based services in widely varying areas such as e-Money and m-Banking, e-Governance, Agriculture, distance Education and m-Health. Therefore, mobile broadband is a strong driver for improving well-being, economic growth (GSMA, 2016, pp.14-30; Broadband Commission, 2016), inclusion (Broadband Commission, 2016), equality and social impact (WDR, 2016). Moreover, mobile broadband helps to lift millions of people out of poverty, having contributed US\$ 3.1 trillion to the global Gross Domestic Product in 2015 alone (Broadband Commission, 2016). One reason for this great success is that the access prices have fallen significantly, according to the ITU Information Society Report 2015. This results in mobile broadband subscriptions being cheaper than fixed broadband ones (ITU, 2015c). Moreover, the same report states that prepaid mobile broadband offers are still the most affordable options for end-users, especially in the remote and rural areas. Recently, India ranked 131st place in the International Telecommunications Union's International Development Index (IDI) 2015.

This report regularly assesses ICT access, ICT usage and ICT skills since 2009. Interestingly, India's ranking (131st) is down six places, compared to five years earlier (ITU, 2015b). Unfortunately, the report provides no reasoning for that downgrade. However, it indicates that other countries developed quicker than India. Nevertheless, this report provides a valuable indication on India's development given their objectives to measure the level and evolution of ICT developments over time, while also highlighting progress in ICT development, the digital divide and further development potential. A major limitation of the ITU International Development Index (IDI) is its lack of metrics for mobile broadband affordability, or infrastructural readiness (see ITU-D, 2017). While being indirectly reflected (e.g. in the number of active mobile-broadband subscriptions per 100 inhabitants), we believe that it would be important that such metrics should be added to the ICT access Index. Moreover, we argue that the ICT usage and the ICT access, as pre-requisite of ICT usage, do not capture the same features from an ICT development perspective. Another relevant report was recently commissioned Facebook's internet.org unit. This represents an Internet inclusiveness report ranking India as the first country for having appropriate policies to ensure future connectivity in place (EIU, 2017). India achieves this mainly due to their July 2016 established 'Aadhar' digital identifier programme and to the recent INR 100 billion (\$1.5 billion) investment to connect some 250,000 panchayats (village councils) by the end of 2018. Nevertheless, in the report, India only ranked 36th out of 75, due to their lower rankings in the other categories. India reaches its worst place for the Availability section (Rank 46), which examines the quality and breadth of available infrastructure required for access and Internet usage. Here, the usage (place 55) and quality (place 49) are ranked particularly low. Moreover, India also greatly lags behind in the Affordability category (Rank 26), which examines the access cost relative to income and competition in the Internet market. In detail, India ranks a good 22nd place in the competition section but a very low 51st rank in the price section (EIU, 2017), a key metric in this dissertation. This rank supports an argument by the International Telecommunications Union, which suggests that a 500MB mobile broadband price plan represents a large portion of the average income (ITU, 2015c) of those Indians who fall under the World Bank's poverty income measurement of US\$3.10/day. This also reflects the urban-rural disparities between the Tamil Nadu districts, as described above.

Moreover, the 2015 Measuring the Information Society Report (ITU, 2015c) display the Indian mobile-cellular sub-basket (monthly cost of prepaid low-user including voice and

SMS services) for 2014, which shows that mobile broadband absorbs 2.14% of Gross National Income per capita, placing India at the 89th position in the world rankings for this criterion ITU (2015c, p.102). In the fixed broadband sub-basket, comparing all available price data of fixed broadband subscriptions, India ranks 108th with a 5.28 % of Gross National Income per capita necessary to afford the basket, while prices are 14 times less affordable in developing countries compared to developed ones. For the mobile broadband (post-paid at 500MB) sub-basket, India ranks 97th with 2.51% of Gross National Income per capita (ITU, 2015c). Hence, affordable and equal access to mobile broadband remains a key challenge, as the Broadband Commission (2014, p.3) states. Moreover, India's income inequalities result in a decline in the country's Affordability Drivers Index (ADI) from rank 30 in 2014/2015 to rank 31 in the 2015/2016 (A4AI, 2016) and rank 35 in 2017 (A4AI, 2017, p.10). While the methodology of this report captures infrastructural aspects, e.g. the extend of ICT infrastructure deployments and policies (see A4AI, 2017), it still lacks other infrastructural metrics such as those reflecting upstream Internet market structures, a crucial aspect of the supply-side of infrastructural access. Nevertheless, the downward trend of India's ranking in the ADI is also apparent in The Web Index of the World Wide Web Foundation (2014). Here, India reaches a low 58th rank for Access and Affordability (Rank 48 for the cost of mobile broadband per capita income) in 2014 and 2017, while also stating that India has no effective law and regulations for Net Neutrality in place. Furthermore, they state that India shows evidence of discriminating practices in practical violation of Net Neutrality criteria (World Wide Web Foundation, 2017). However, the latest Global Information Technology Report's Network Readiness Index (NRI) by the World Economic Forum conflicts with the other indications (WEF, 2017a). Here, India reaches rank 8 out of 139 in terms of affordability (WEF, 2017, p.110; 2017b). This rank has to be considered with great caution since it does not include mobile broadband in the affordability section. As already shown by Ericsson (2016), improving affordability is expected to lift the Indian mobile broadband penetration rate to 68% by the end of 2020, which represents a staggering increase of 330 million new mobile broadband subscribers (GSMA, 2016). These users are most likely upgrading from their existing non-broadband cellular subscriptions. Nevertheless, Onno Ruhl, the World Bank Country Director India states that:

'However, to reap the full benefits will require affordable and wider access to the internet and skills that enable all workers to leverage the digital economy.' (The World Bank, 2016c).

Moreover, affordable subscriptions would help to increase India's rank in the ICT Development Index (IDI), especially after the country's recent drop by 3 places compared to 2015 (Rank 138 in 2016). More affordable price plans would also directly affect the ICT Development Access and Use sub-indexes of the International Development Index (ITU-IDI, 2017). The Access sub-index captures the ICT readiness of a country and therefore includes five infrastructure indicators (fixed telephone subscriptions, mobile-cellular subscriptions, international Internet bandwidth per Internet user, households with a computer and households with an Internet access). The Use sub-index captures the intensity of ICT, covered by three indicators (individuals using the Internet, fixed broadband subscriptions and mobile broadband subscriptions). India could furthermore increase their position in some of the metrics of the Network Readiness Index (91st rank out of 139 nations) by the WEF (2017a). Moreover, affordable price plans for end-users could increase access to the mobile Internet for the poorest strata of the population, helping to drive the Sustainable Development Goals (SDG) of the United Nations (2017).

Unsurprisingly, studies on mobile broadband are of particular relevance, according to the World Bank Digital Dividends Background Paper. Minges (2016) states that mobile broadband is especially important given its rapid diffusion in developing countries. Estimating a panel data model for a sample of developed countries covering the period 2005-2009, Thompson and Garbacz (2011) show that a 10% increase in the 3G penetration rate raises the annual GDP growth rate by 0.15%. The doubling of mobile data consumption across the 14 countries studied is considered to have raised GDP by 0.5% (Thompson and Garbacz, 2011). Hence, mobile broadband penetration and its usage are found to boost economic growth. To exploit these opportunities for economic growth through mobile broadband, the Government of India is currently embarking on the implementation of 'Digital India' (Broadband Commission, 2016), a rural programme to connect 2.5 million panchayats, or village councils, as indicated above. Moreover, the mobile broadband operators are greatly investing in strengthening their 3G and 4G network coverage.

Besides the economic impact on GDP, the strengthening and diffusion of the mobile technologies have additional social impacts. Unfortunately, the measurement of social impacts of Information and Communication Technologies tends to attract less attention than that of the economic ones (OECD, 2008). The International Telecommunications Union considers broadband as a tool for poverty reduction and consumer welfare (ITU,

2014, pp.2-35). Moreover, they argue that mobile broadband is the primary platform for emerging markets. In a classification among *constrained*, *emerging*, *transitional* and *advanced* stages of development, India has been ranked as being in the *emerging* category. This represents a stage where mobile broadband is considered to be a strong enabler for positive social impact (Katz, Koutroumpis and Callorda, 2013). Interestingly, these authors also state that broadband has been found to increase monthly income, which should usually be the reverse, for a sample in Ecuador (Katz, Koutroumpis and Callorda, 2013, p.34), the overall effect seems greater for men than for the woman, therefore widening the gender gap. This issue is also considered by the United Nations (UN, 2014) and Fennell and Arnot (2007), who ascribe it to the gender differences in education, lack of income, and social attitudes (towards technology). According to Katz, Koutroumpis and Callorda (2013), this gap arguably disappears when the broadband users were previously Internet users, compared to those who use the Internet the first time through mobile broadband. However, the affordability issue related to gender disparities in income remains. In an Australian survey, Siddhartha De (2007) reveals that Information and Communication Technologies (ICT) impact many facets of people's everyday life. These are mostly related to the accessing of, and interaction with, information as well as communicative relationships with family and community members. The OECD (2008) also supports this evidence, stating that mobile broadband brings social benefits such as social connection amongst consumers, connections to businesses and governments. According to Van Dijk (2006), the term 'digital divide' causes more confusion than clarification but refers to, amongst other effects, the uneven development of the Internet throughout the world. Selvabaskar et al. (2016) refer to the 'digital divide' as an obstacle to use ICT and propose a separation to tele-density, mobile and Internet divide as introduced by Parvathamma (2003). Furthermore, they note that the unequal development arises due to rising population, inadequate funds, affordability issues and policy implementations thereof (Selvabaskar et al., 2016, p.2). Moreover, the Mosaic Group (1998) developed the Internet Diffusion Framework as used by researchers such as Castells (2001) or Rogers (2001). This Framework serves as a methodology for analysing the state of a country's Internet diffusion. Guillén and Suárez (2005) find that the number of Internet users is a widely used indicator of the level of a country's development. Considering that the number of Internet users is often not reliable, as shown before, its utilisation as a development index seems rather surprising to us. A number of researchers such as Bagchi (2005) and Deichmann et al. (2007) find that the numbers of computers

per 100 people are correlated with GDP per capita, a metric that seems more pragmatic. On the other hand, Shea, Ariguzo and White (2006) argue that often even in the presence of access, only a small number of the global population benefits from using the Internet. Paul (2002) states that the gaps as well as the development in Information and Communication Technologies are not static but that the digital divide refers to an unequal and disproportionate pace of development in access to digital infrastructures and services (Paul, 2002, p.13). In a study on the diffusion of the Internet in India, using the Global Diffusion of the Internet (GDI) framework, Wolcott et al. (2001, p.17-33; 2001) state the importance of connectivity and organisational infrastructure, which shows a lack of a user-centric approach towards collecting reliable information on connectivity infrastructure measurements. Wolcott (2005) relies on the number of Indian *Internet Service Provider* license holders and their points of presence, amongst other indicators, to study the Indian policy landscape. However, we argue that the sheer presence of an *Internet Service Provider* in a region does not necessarily imply their connectivity quality or the underlying market pricing for both the *Internet Service Providers* and the affordability of Internet subscriptions for end-users. Selvabaskar et al. (2016) point out that the diffusion of mobile technologies is not uniform across Tamil Nadu. Other researchers consider additional factors for the ‘digital divide’ such as public and private initiatives towards Information Technology education, science and technology investments, the cost and regulations of *Internet Service Provider* services (Bennett and Norris, 2001), or the Internet and broadband (Kagami, Tsuji and Giovannetti, 2004). The latest World Bank World Development Report (WDR, 2016, p.4) defines the ‘digital divide’ more appropriately as making the Internet more accessible, affordable, open and safe. Hence, closing the ‘digital divide’ would result in spreading benefits and reducing risks of the so-called ‘digital dividends’. The benefits include reaching new services, increased efficiency for accessing affordable activities and services, and innovation due to lowering transaction costs. Making the Internet more accessible, open and safe for Indians must be a vital priority to close the ‘digital divide’, while strengthening regulations that ensure competition among businesses and accountability of governmental oversight (WDR, 2016). The risks of the ‘digital dividends’ refer to governments and corporations controlling citizens, inequality due to labour markets and a harmful concentration of economic sectors. The ‘digital divide’ on the openness of the Internet often refers to ‘Net Neutrality’, which the WDR (2016) itself considers as a confounding issue. ‘Net Neutrality’ refers to a fair treatment of all data that is travelling through the

networks of *Internet Service Providers*, without discriminating in favour of particular websites or services, according to the Electronic Frontier Foundation (2016). Therefore, a network is considered neutral when *Internet Service Providers* manage all types of data packets equally. The EFF states that *ISPs* should not become gatekeepers for special deals, which would inhibit competition, innovation and freedom of expression (Electronic Frontier Foundation, 2016). However, the discussion ranges around the management and prioritisation of scarce bandwidth resources, mainly driven by content providers. According to Choi, Galeotti and Goyal (2014, p.3), policymakers aim to quantify network market power within the ‘Net Neutrality’ debate. While standard metrics fall short, they argue that sophisticated metrics based on the structure of the Internet are currently in need of development (Choi, Galeotti and Goyal, 2014). This agrees with the WDR (2016), which states that ‘Net Neutrality’ might refer to discussing resources or free speech. Moreover, the report refers to freedom of expression and access to information and therefore human rights issues. Facebook’s Internet.org ‘Free Basics’, ‘Wikipedia Zero’ and Free access to Facebook and WhatsApp by Aircel, free access to Google by Bharti Airtel, and free access to Twitter by Reliance provide examples of recent violations of the ‘Net Neutrality’ principles in the mobile broadband markets in India. Such examples forced the Telecom Regulatory Authority of India to set preliminary ‘Net Neutrality’ rules in their Prohibition of Discriminatory Tariffs for Data Services Regulations in February 2016 (TRAI, 2016e). Prior to that, as of August 2015, the Indian government had released no policy statements on ‘Net Neutrality’ (The Editorial Board, 2015). Nevertheless, an open and free Internet is considered as a key factor for innovation and inclusion in digital economies, where users should have equal and affordable access to the Internet, its content and services. However, traffic management on the Internet is legitimate and should not reduce the fundamental rights and freedoms of end-users (WDR, 2016). Moreover, the United Nation argues that the balance between a free traffic routing choice of *Internet Service Providers* and the accessibility and freedom of end-users should be properly balanced to continuously incentivise improvements in networks. The recent developments from the Federal Communication Commission (FCC) in the United States tilted the Open Internet Order (OIO) rectifying *Internet Service Providers* with the power to control access and content distribution (Forbes, 2017). More recently, the FCC, under the direction of US president Donald Trump, repealed the Net Neutrality law in the United States. This gives large access providers more power and control without regulatory oversight, while risking worsening affordability of the Internet access.

Finally, to explore the access of end-users to affordable price plans and the power of *Internet Service Providers* on the upstream Internet market, we embark, in the next section, in the review of these issues from a Network Science perspective. We consider this approach as particularly relevant since it provides the power to capture key underlying economic relationships among the different interacting providers on the Internet.

2.3 Internet Market Structuring

The use of Network Science provides additional measurements of valuable relationships among interacting economic providers (Schneider and Bauer, 2016, p. 73), particularly when embedded in real-world systems (Albert and Barabási, 2002). The fields where Network Science methods have been applied are extremely diverse, including the analysis of movie actor networks (Watts and Strogatz, 1998), Computer Networks (Wasserman and Galaskiewicz, 1994), Social Networks (Wasserman and Faust, 1994; Scott, Carrington and Weihua, 2011), and Bioengineering (Schuster et al., 2002), amongst many others.

The interactions amongst economic providers to transfer digital goods and services through data and communication flows over the Internet can also be studied using Network Science. Early approaches such as Alvarez-Hamelin et al. (2008) applied Network Science for studying upstream Internet connectivity but they do not focus on the economic side of this network infrastructure. The collection of transaction data among economic agents on the Internet was previously considered as expensive and arduous (Newman, Barabási and Watts, 2006; Schneider and Bauer, 2016). However, this is true only when considering traditional and non-creative solutions to this problem, as collecting and analysing such data, as it is done in this dissertation, is becoming increasingly easier through data collection via active and passive Internet measurements.

While Schneider and Bauer, (2016) refer to the rising affordability of storage and computing power, we see new data collection methods, such as crowdsourcing, as a key benefit. However, the emerging research area at the interface between Network Science and Internet Economics is still in an early stage of development (Schneider and Bauer, 2016). This shows the lack of valuable interdisciplinary research where we aim to engage in. Traditionally, the literature on the economics of networks focuses on the functioning, formation and structuring of networks (Vega-Redondo, 2003; Jackson, 2008; Goyal,

2009), mainly from physics, mathematics or economics standpoints. The following sections will explore the collection of Internet connectivity data before looking at various areas that study the relevant network formation and structuring approaches required to explore this dissertation's research questions.

2.3.1 Collection of Network Data

Due to its heterogeneity and increasing complexity, the Internet as a whole system is considered immeasurable (Murray and Claffy, 2001). Nevertheless, measuring some key structural properties of the Internet may be done by following active or passive measurement approaches. The most common methods to date are passive measurements usually built into routers or switches that track Internet traffic as it is routed through them. On the contrary, active measurements inject test data packets into the networks to 'sniff out' responding information from devices such as routers (SLAC, 2001).

Most of both active and passive measurement techniques utilise crowdsourcing approaches. Single measurements for specific Internet problems such as structural bottlenecks analysis rather than complete Internet mapping efforts is considered especially valuable (Murray and Claffy, 2001). However, we feel that their value lies not only in the identification but also in the rerouting of traffic around the discovered bottlenecks.

The Internet community and applications saw great examples of crowdsourcing efforts such as the collection of crisis information by Ushahidi (2016) or the 2001 launch of Wikipedia (Wikimedia Foundation, 2016). The term 'crowdsourcing' was first coined by Jeff Howe in a Wired (2006) magazine article and refers to a

'participative online activity in which an ...institution, ...or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, ...the voluntary undertaking of a task.'

Comparing 175 research articles, Estellés-Arolas and Gonzáles-Ladrón-de-Guevara (2012, p.11) generate a conclusive crowdsourcing definition, which mostly agrees with the one proposed by Jeff Howe. Numerous efforts aim to measure and characterize the structural properties of the Internet in using crowdsourcing measurements. A crucial work can be found in the research of the DIMES project, studying the Internet structure with the help of a voluntary community (DIMES, 2012). The software agents that DIMES

distribute, which autonomously run software programs installed on privately owned machines, measure connection *traceroutes* and their ping (see Glossary) times for diagnosis purposes. *Traceroutes* refer to a network diagnostics and measurement tool, using the Paris Traceroute (2016) version, to display and measure the paths of data packets across Internet Protocol networks. The last activities of DIMES (2012) seem to date back to the year 2012. Nevertheless, through their crowdsourcing of agents, DIMES demonstrate the ability to discover hidden parts of the Internet structure, as Shavitt and Weinsberg (2011) valuably show. Nevertheless, the DIMES project did not link their data findings to end-user affordability of access from the Internet periphery. Another research project that is making use of crowdsourced agents is the 2007 launched CAIDA Archipelago, or CAIDA-Ark (2016), building upon their previously 2008 retired Macroscopic Topology Project ‘Skitter’ (see CAIDA, 2016c)) and the DIMES project, where networks can participate by hosting so-called ‘Ark monitors’ to collaborate specifically towards active network measurements of the Internet structure. While the CAIDA-Ark (2016) aims to focus on Internet Topology Discovery and Congestion, key economics questions and incentives are also left out. The still-active RIPE NCC Atlas project follows a slightly different crowdsourcing approach than DIMES since it provides a testing infrastructure for community members, mostly network providers like in the CAIDA Archipelago, being interested in performing Internet connectivity and reachability measurements through a hardware probe (RIPE NCC, 2016). Therefore, the RIPE NCC Atlas follows the CAIDA-Ark (2016) best practices. Just like DIMES, the Atlas also makes use of *traceroutes*, amongst other technologies. The RIPE NCC Atlas represents a successful crowdsourcing project with a large number of 9,334 connected probing devices as of October 2016 (RIPE NCC, 2016b). Here, again, RIPE focusses on real-time Internet usage measurements rather than on the key economic dimensions of the Internet structure.

When it comes to Internet crowdsourcing projects that involve end-users rather than network providers, there are currently four notable projects in the research environment, Netalyzr, Netradar, OpenSignal and Portolan. First, the UC Berkeley originating ICSI Netalyzr is primarily a debugging tool for testing network connection issues on Google Android devices (ICSI, 2016). The Berkeley International Computer Science Institute (ICSI) uses the Netalyzr tool for network diagnostics, measuring the health of the Internet’s edges, rather than the structural properties of it. Both Netradar (2016) and OpenSignal (2016) are providing Android and iOS applications for measuring the signal

coverage and performance of mobile broadband operator networks that can be embedded in Google Maps. While Netradar (2016) comes from Aalto University's School of Engineering, OpenSignal (2016) is provided by a London-based venture-backed company. Lastly, the Portolan represents a joint research effort of the Istituto di Informatica e Telematica of the Italian National Research Council CNR (IIT) and the University of Pisa (Portolan, 2015). Portolan is an active Internet measurement project and tool that aims to discover the structure of the Internet as well as signal coverage maps similar to Netradar and OpenSignal. The Portolan Project also relies on crowdsourcing of data collection through an application for Android end-user devices. The Portolan Project is a unique approach to measure the Internet structure from an active Internet periphery perspective, as indicated by Faggiani et al. (2012; 2013; 2014a; 2014b) and Gregori et al. (2013). This allows the study of mobile broadband operators' Autonomous Systems from a unique end-user perspective, as shown in a pilot case study on a Bhutanese incumbent mobile broadband operator by Giovannetti and Sigloch (2015). Some researchers from the computer science field (e.g. Vázquez, Pastor-Satorras and Vespignani, (2002)) argue that *traceroutes* analysis at Internet Protocol level from one location in the network are unreliable when constructing complete Internet mapping projects, due to cross-links and other technical issues. Knight et al. (2011) mention that *traceroutes* are commonly used but also point towards the deficiencies of such measurements, mainly supporting the work of Willinger, Alderson and Doyle (2009). However, Knight et al. (2011) also point towards the possibility of analysing networks at Autonomous System granularity. Other researchers such as Feldman and Shavitt (2008), Siganos et al. (2003), Alvarez-Hamelin et al. (2008) and Giovannetti and Sigloch (2015) transform IP addresses to the AS granularity in order to reveal a greater understanding of the upstream Internet market structure. Such transformations, however, are highly dependent on the reliability of secondary fusion datasets, a major downturn. In her early work, Gao (2001) considers the Autonomous System level as especially valuable for analysing commercial contract relationships amongst *Internet Service Providers*. Dimitropoulos et al. (2013) agree on this and consider the AS granularity especially valuable when merged with a secondary CAIDA (2016b) AS-relationship dataset. Unfortunately, we believe that those commercial contract relationships are not entirely discoverable in practice, given more informal business relationships between Autonomous Systems.

Insight 5: Due to its applicability to measure the upstream Internet market structure, from an Internet periphery perspective, the Portolan (2015) application seems more appropriate compared to the alternatives Netalyzer, Netradar and OpenSignal that focus on different infrastructural specificities. By additionally using, filtering and integrating secondary data, obtained from Maxmind (2015), we are able to analyse IP and AS granularities to understand the upstream Internet market structure in our case studies. In doing so, we follow best practices of the Computer Science researchers such as Alvarez-Hamelin et al. (2008). However, in addition to these researchers' contributions, we also add a critically relevant end-user perspective, focussing on the conditions of accessing the Internet infrastructure from these networks' periphery.

2.3.2 Network Formation

Network Models

Network Science generally studies the forces that shape developments of networks and their structuring. Networks are composed of vertices (representing network's agents) and their edges, linking them, representing relationships between these network vertices. According to Schneider and Bauer (2016), empirical networks on the Internet are considered to be neither regular nor random. Regular networks refer to network graphs where each vertex has the same number of neighbouring vertices and every vertex has the same number of *In-* and *Out-Degrees*, representing incoming and outgoing relationships between vertices. Networks may be studied based on their edges being of a directed or undirected nature. Directed networks refer to relationships amongst edges that are directed (e.g. vertex A links to vertex B but vertex B not to vertex A). Undirected networks merely acknowledge if there is a linkage (and possibly its number of occurrences) between vertices, or not. The literature covers a number of network formation models that follow specific structural properties. *Random Networks* employ probability distributions (Bollobás, 2001) and were first defined by Erdős and Rényi (1959) and independently by Gilbert (1959). Watts and Strogatz (1998) and Watts (1999) propose a *Small-World Network* model where the vertices in so-called sub-graphs (subsections of network graphs) are densely interconnected amongst each other. Watts and Strogatz (1998) are credited for this model but base their work on earlier models of Simon (1962). Albert and Barabási (2002) introduce the preferential attachment ('rich-get-richer') effect of vertices and edges in so-called *Scale-Free Network* models. These preferential attachment models allow researchers to simulate the emergence of growth in

networks, which are discussed as network effects and network externalities in economics, where Katz and Shapiro (1985) and Economides (1995) amongst others, discuss the implications of network externalities on the telecommunications market structures.

Power-law Degree Distributions

By looking at the aggregate properties of the resulting distributions, the preferential attachment modality of establishing connections between vertices in a network leads to the presence of *power-law degree distributions* (Albert and Barabási, 2002; Barabási Labs, 2013). These *power-law degree distributions* are typical indicators of the presence of a hierarchical network structuring since a few vertices have many edges directly linking them with other vertices, whereas many vertices only have a few edges. This feature is typically captured in distributions of edges, following *power-law degree distributions* (Pareto, 1906) which have also proven useful in modelling income distributions (Reed, 2001). Faloutsos, Faloutsos and Faloutsos (1999) find that the Internet structure follows *power-law degree distributions* at the Autonomous System level. The work of Dall'Asta et al. (2005) supports this finding. When comparing different tools for generating network structures, Medina, Matta and Byers (2000) argue that *power-laws* can only be found in dynamical growth models such as the one of Barabási and Albert (1999), which adds new vertices and edges to a network. Hence, Medina, Matta and Byers (2000) provide sufficient proof that outgoing connectivity of a vertex (see section 3.4.2 below) and rank exponents (preferential attachment of edges and vertex growth (Barabási and Albert, 1999)), provide 'useful means' for testing the structure of the Internet. Before that, Crovella and Bestavros (1996) find that the Internet at the World Wide Web level also displays *power-law degree distributions*. This is supported by the findings of Albert, Jeong and Barabási (1999), Huberman and Adamic (1999) as well as Kumar et al. (1999). Caldarelli, Marchetti and Pietronero (2000) then find, on the basis of Internet mapping efforts by Cheswick, Burch and Branigan (2000), that an analysis at Router-Level (Internet Protocol granularity) from an end-user perspective, also shows *power-law degree distributions* as well as *Scale-Free Network* properties. While Pastor-Satorras, Vázquez and Vespignani (2001) and subsequently Vega-Redondo (2003) support these findings, Knight et al. (2011) argue that *power-law degree distributions* seem convincing but lack accurate data, since the data used was not published in line with the usual articles. Lakhina et al. (2003) also argue against Faloutsos, Faloutsos and Faloutsos (1999), saying that *power-law* functions are an illusion of biased data. More recently, Willinger and Roughan (2013) also challenge the *power-law* analysis saying

that *traceroutes* detections at Internet Protocol (IP) Level are representing network specifics (opaque layer 2 cloud networks) and add that *traceroutes* are unable to reveal the actual vertex degree of any routers. They conclude that the absence or presence of *power-law degree distributions* cannot be justified with reasonable statistical confidence. While taking a pragmatic stance on this issue, we recognise the importance to choose the most appropriate granularity of analysis.

Insight 6: Based on the research findings stated in the above literature, we expect our case study networks to display *power-law degree distributions* for primary collected active Internet periphery measurements. Given the economic nature of the upstream Internet market, these *power-law degree distributions* are signalling the presence of a Tier-Model of Internet Service Provider relationships (see e.g. Luckie et al., 2013), and can be used to explore the presence of hierarchical structuring in the upstream Internet access market.

Levels, or Granularities of Analysis

Research in the Computer Sciences does not seem to be reaching a consensus on the most appropriate level of granularity for the analysis of Internet networks. Faloutsos, Faloutsos and Faloutsos (1999) mention two possible levels of analysis, namely the Router level (Internet Protocol) and the Inter-Domain Level (Autonomous Systems). Vega-Redondo (2003) agrees on these two levels of analysis. Others such as Huffaker, Fomenkov and Claffy (2016) from CAIDA define six possible granularities of analysis, namely the Fiber, IP address, Router, Points-of-Presence, Autonomous System and *Internet Service Provider*. Just like Faloutsos, Faloutsos and Faloutsos (1999), Willinger and Roughan (2013) mentioned the Router level but elaborated further on the Switch granularity (IP Links between hubs and switches), the Physical level (including all Layer 1 devices), the Point-of-Presence Level, the Application Layer such as HTTP and HTML and finally the Autonomous System Level. When analysing the economics of Internet routes, Kagami, Tsuji and Giovannetti (2004) differentiate between three layers of analysis: the end-user level, the *Internet Service Provider* level and the major Internet backbone providers. These layers can be divided into the traditional supply and demand sides in economics. This represents a valuable departure from the more technical approaches of Computer Science. Given all these levels of granularities, a thorough structural analysis becomes impossible, since one might study links and flows between physical objects as well as information (Willinger and Roughan, 2013). By analysing the difficulties of simulating

the Internet, Floyd and Paxson (2001) also reveal the great heterogeneity for studying the individual links of network traffic or the information flow through protocols on top and argue that the structure of the Internet is difficult to characterise due to its ever-changing dynamics. Nevertheless, the study of the Internet structure at different granularities, especially the Router and Autonomous System granularities, is considered to be of equal and fruitful importance (Faloutsos, Faloutsos and Faloutsos, 1999). Moreover, the results may represent a Complex Network architecture composed of many vertices and few relationships amongst the vertices (Vega-Redondo, 2003). Gorman and Malecki (2000) argue, that a combination of Network Analysis for studying the Internet structure is a surprisingly under-researched field. However, one has to choose the most appropriate granularity of analysis, given the research problem at hand.

Insight 7: Discussing the different granularity assures us that our exploration should be most valuable using the Internet Protocol and Autonomous System granularities, following best practices of the early Computer Science literature such as Faloutsos, Faloutsos and Faloutsos, (1999). Moreover, we believe that the Autonomous System granularity allows us to shed light on economic relationships amongst Internet Service Providers.

Detailed research that relates to our case study is very limited and mainly attributes to the following research papers. In using four different datasets, Barnett and Park (2012) investigate the structure of the World Wide Web (WWW) using Network Analysis. Their findings indicate that the Internet consists of a series of *Small-World Networks*, which only seems applicable for WWW networks. However, fully interconnected sub-graphs at IP or Autonomous System granularity appear counter-intuitive, given the connectivity role played by Tier-1 *Internet Service Providers* as described in section 2.1.2 above. More recently, and most related to this dissertation, is the work of UC Davis researchers Ruiz and Barnett (2015), who study the International *Internet Service Provider* (ISP) ownership network at company and national levels. Their approach relies on secondary Telegeography Autonomous Systems data for 113 companies and captures the number of *Internet Service Provider* relationships, their vertex degrees as well as the *Eigenvector* and *Betweenness Centralities*. The findings of Ruiz and Barnett (2015) show that Level 3 Communications, Century Link, Telia Sonera, AT&T and Cogent Communications are the most central companies in their limited dataset. This finding was to be expected, given the role of the Tier-1 *Internet Service Providers* that their study finds as well as the

CAIDA (2016a) AS-Rank data. However, their study fails to employ additional relevant network metrics and therefore lacks an in-depth analysis of structural network phenomena. Moreover, Ruiz and Barnett (2015)'s work is based on secondary data rather than primary collected active Internet periphery measurements, representing an end-user access to the upstream Internet market. Also, closely related to this work is the pilot case experiment by Giovannetti and Sigloch (2015) who study the incumbent Bhutanese Mobile Broadband operator network at IP and AS granularity using active Internet periphery measurements. Their analysis of primary active Internet periphery measurements using the *Clustering Coefficient* metric reveals the structural properties of the upstream Internet market, while also indicating previously hidden upstream Autonomous System relationships. While this pilot work opens up an entirely new field of research, they also lack to link it to end-user affordability.

Insight 8: Giovannetti and Sigloch (2015) find previously hidden Autonomous System relationships that were not visible in the CAIDA (2016b) dataset. These hidden relationships were identified using the *traceroute* analysis at IP and AS granularity for a Bhutanese mobile broadband operator. We infer that an extension of the employed analytical approaches should also reveal hidden AS-relationships for the mobile broadband operators in this case study (see below).

2.3.3 Network Structuring

Once appropriate granularities of analysis are identified, Network Science provides useful metrics to study relevant structural network properties to unravel economically interesting connections amongst *Internet Service Providers*. Such metrics focus on capturing network features such as densities, centralities and clustering to explain network formation and structuring. Ever since the introduction of Social Network Analysis, researchers applied a multitude of centrality metrics for studying the importance of specific network agents and relating them to structural positioning (Wasserman and Faust, 1994). The *Clustering Coefficient* in a directed graph provides a ratio between existing edges, amongst all other vertices being connected to the same vertex, over the number of possible interconnections (Boccaletti et al., 2016). This metric is useful to explore the bargaining power of central agents as Vázquez, Pastor-Satorras and Vespignani (2002) point out from a more technical perspective. Moreover, this metric is used by Giovannetti and Sigloch (2015) to study the upstream Internet market structure for a Bhutanese incumbent mobile broadband provider. Sociological research such as

Cook et al. (1983) consider structural network centrality as one of the most relevant properties to study. Early work by Freeman (1979) identifies metrics to study communication activity between network vertices, namely *Degree Centrality*, *Betweenness Centrality* and *Closeness Centrality*. Freeman (1977; 1978) assumes that the *Betweenness Centrality* of a vertex is suitable to measure the resource control of a vertex over other network vertices. Hence, vertices with a high *Betweenness Centrality* should govern, according to Freeman (1978), a strong influence on the transfer of goods in a given network. Nevertheless, the *Betweenness Centrality* encapsulates the assumption that the transfer of goods follows the shortest paths in a network (shortest distance between vertices), while we refer to the transfer of digital goods as a commodity service in the digital supply chain. This does not necessarily apply for connectivity on the global Internet infrastructure. Moreover, Cook et al. (1983) note that vertices with a high *Betweenness Centrality* should exert a strong control on the network. Bonacich (1987) from the UC Los Angeles states that more central vertices have greater power on the connectivity to other vertices, determined by the number of central vertices it is connected to, and therefore a strong network influence. Stephenson and Zelen (1989) additionally note that more central vertices also receive more information. Contrary to the Social Science researchers, Network Science researchers from Physics and Mathematics including Mintz and Schwartz (1985), Monge and Contractor (2003), and Knoke and Yang (2008) argue that a vertex' power does not equal to its centrality in a network. Additionally, Stephenson and Zelen (1989) also argue against the *Betweenness Centrality* of Freeman (1977) and suggest that communications between people does not necessarily follow shortest paths. Again, we expand here and assume that internetworking connectivity on the Internet also does not necessarily follow shortest paths, given the underlying business relationships between *Internet Service Providers*. Freeman, Borgatti and White (1991) incorporate this feedback by proposing the *Flow Betweenness Centrality*, a measurement that takes the weights of relationships into consideration. Due to the *Betweenness Centrality*'s issue of requiring long calculation times, Brandes (2001) finds a more efficient calculation algorithm, which is used in the work of D'Ignazio and Giovannetti (2006). Again, given the issue of shortest paths, Noh and Rieger (2004) find that the *Betweenness Centrality* is only appropriate when the global connectivity of each vertex is known and propose the *Random-Walk Centrality*, a metric for which only local vertex connectivity is known. Nevertheless, none of these *Betweenness*-based metrics necessarily represents real-world networks. Newman (2005) incorporates the idea of the

Random Walk Centrality and the *Betweenness Centrality* into the *Random Walk Betweenness Centrality*, while providing some examples of scientist networks and sexual contacts. However, no variation of the *Betweenness Centrality* sufficiently solved its shortcomings to date, except for adding great complexity. More simple vertex metrics are Freeman (1979)'s *Degree-* and *Closeness Centrality*. Here, the *Degree Centrality* simply measures a vertex importance given its number of connections, as Newman (2006) also points out. However, the *Degree Centrality* does not take the direction of an edge into consideration and poses restrictions on directed graphs. This means, that the applicability to study connectivity flows on the upstream Internet market is fairly limited. Moreover, this results in the separation of the metric to *In-Degree Centrality* for incoming connections and *Out-Degree Centrality* for outgoing ones. The third metric of Freeman (1979), *Closeness Centrality*, simply defines the theoretical distance of a vertex from all other network vertices. From another perspective, Bonacich (1972) proposes the *Eigenvector Centrality* as an alternative metric. The *Eigenvector Centrality* assigns relative scores to vertices, under the assumptions that vertex connections to high-scoring vertices in the network contribute more to its score than equal connections to low-scoring vertices, and as such, it measures the overall influence of a vertex in a network (Bonacich, 1987).

Therefore, the *Eigenvector Centrality* is not only suitable for graphs with strictly binary vertex relationships, such as *Degree Centrality*, *Closeness Centrality* or *Betweenness Centrality*, but also to those with less trivial relationships (Bonacich, 2007). Hence, the *Eigenvector Centrality* is particularly suitable for networks that employ vertices with high degree positions connected to many low degree vertices, and vice versa. The metric could, therefore, capture *power-law degree distribution* structures as well as situations with less connected network peripheries. Moreover, this metric could be particularly useful when considering the real-world nature of recurring business relationships. Ruiz and Barnett (2015) use, amongst other metrics, the *Eigenvector Centrality* to identify 'central' companies using commercial Telegeography data. Another important, but less widely adopted centrality metric, is given by the *Katz Centrality*, which measures a vertex influence given its total number of walks between a pair of vertices. According to Newman (2010), the *Katz Centrality* is better suited to analysing so-called directed acyclic graphs than the *Eigenvector Centrality*, where connections between vertices can only take one direction, from the periphery to the network core. These situations seem to be apparent in citation networks or the World Wide Web but do not fit to *traceroutes*,

where destinations may be situated closer to the Internet edge (Newman, 2010), while still passing by the core of the Internet, where necessary.

Insight 9: Linking the theoretical understanding of *traceroute* data to the above centrality metrics discussion indicates that the greatest value for an economic analysis should be found in using the *Clustering Coefficient* (see Vázquez, Pastor-Satorras and Vespignani, 2002 and Giovannetti and Sigloch, 2015) and the *Eigenvector Centrality*. Especially the latter metric has been highly neglected and was not used before to capture hierarchical structuring of upstream Internet markets. Only the work of Ruiz and Barnett (2015) uses this metric albeit in a different setting. All other (centrality) metrics seem to suffer from shortcomings such as being limited to only certain network types or assuming prespecified features of connection amongst vertices, e.g. following shortest paths or specific walks in a given network.

The literature that studies criticality of network agents using Network Analysis, including the metrics discussed above, reveals some additional insights. By studying strategic interaction games over networks, Bramoullé, Kranton and D'Amours (2014) solve Nash equilibria showing that all equilibrium solutions are characterised by a players' Bonachich' *Eigenvector Centrality*. Ballester, Calvó-Armengol, and Zenou (2006) find that network centrality relates to games with linear externalities of the network structure when studying investment levels. Other economists such as Galeotti and Goyal (2010) study access of information through network influencers establishing the law of the few, where a majority of individuals get information from the few influencing ones. Golub and Jackson (2010) explain the structure of diffusion of information on the Internet using a network of 10,000 nodes by Liben-Nowell and Kleinberg (2008). Elliot and Golub (2013) study the problem of key network agents' outcomes in public goods cooperation using *Eigenvector Centrality* measures. Others, such as Fershtman and Gandal (2011), focus on the vertex centrality in relation to the diffusion of knowledge spillovers. Banerjee, Chandrasekhar, Duflo, and Jackson (2013) utilise centrality metrics for assessing the diffusion of information as a success element for microfinance loan programmes in 43 Indian villages. In terms of network structuring, Bramoullé, Kranton and D'Amours (2014) introduce the use of the lowest eigenvalue, a graph spectral analysis metric that is seldom used in Network analytical settings. Moreover, they relate this metric to the bi-partitioning of a network graph, meaning that economic agents are divided into two distinct sets of relationships where an agent's actions possibly rebound between the

different relationship sets in the network graph. More closely related to the study of Internet market structures, D'Ignazio and Giovannetti (2006; 2009) study *Internet Service Provider* market concentration in the upstream Internet layers, using secondary peering data from Internet Exchange Points (IXP) while linking them with CAIDA's Customer Cone data. Based on the efficient *Betweenness Centrality* algorithm of Brandes (2001) and early structural parameter calculations of Shimmel (1953), D'Ignazio and Giovannetti (2006; 2009) show that the *Betweenness Centrality* of an *Internet Service Provider* can be used to calculate the traditional *Lerner Index*, an index capturing an organisation's degree of market power. The work of D'Ignazio and Giovannetti (2006; 2009) not only influenced the Computer Scientists Luckie et al. (2013) at CAIDA but also informed further economic studies of networks. Here, Choi, Galeotti and Goyal (2014) use D'Ignazio and Giovannetti (2006; 2009)'s *Betweenness Centrality* based market power to provide evidence for game-theoretical equilibrium pricing and division of agent surplus, and hence the functioning of intermediated networks. The findings of Choi, Galeotti and Goyal (2014) also show that an agent's criticality is relevant in defining its underlying market power while affecting distribution, pricing and the efficiency of economic activities in networked markets.

Closely related to our work are the studies of Pastor-Satorras, Vázquez and Vespignani (2001) and Vázquez, Pastor-Satorras and Vespignani (2002, p.5-11). From a structural metrics perspective, by analysing 6,374 Autonomous System connectivity maps for the years between 1997 and 1999, they argue that the (hierarchical) structure of the Internet can be studied by estimating the empirical relationship between the *Clustering Coefficient* and connectivity. Their findings show that when the *Clustering Coefficient* scales as a negative *power-law* function of the connectivity, then the underlying network is characterised by a hierarchical organisation. While these indications are often neglected in economic research, other research fields tend to explore these relationships. Rubinov and Sporns (2010) provide a metric toolkit for exploring neurological brain connectivity through Complex Network Analysis and point towards the connectivity importance of the *Clustering Coefficient*. D'Ignazio and Giovannetti (2014) and Giovannetti and Sigloch (2015) study the empirical relationship between the *Clustering Coefficient* and *Degree* connectivity. D'Ignazio and Giovannetti (2014) focus their work on global supply chain networks that are interconnected to local ones through Internet Exchange Points (IXPs). Giovannetti and Sigloch (2015) also argue, on the basis of significant negative regression coefficients, that the *Clustering Coefficient* differs in their relationship with *In-* and *Out-*

Degree connectivity within their given network but neglected to further explore the structural importance of certain Autonomous Systems.

Insight 10: Given the structuring of Autonomous Systems on the Internet, we logically infer that the emergence of a hierarchical upstream network structuring, resulting from the relation between the *Clustering Coefficient* and *Degree* connectivity (as introduced by Vázquez, Pastor-Satorras and Vespignani, 2002) could also be apparent for the networks of the Tamil Nadu mobile broadband operators. Giovannetti and Sigloch (2015) showed this for B-Mobile, the Bhutanese incumbent mobile broadband operator. Such a hierarchical upstream Internet market structure (following *power-law degree distribution*) indicates that a few large Autonomous Systems provide key connectivity to the many smaller ones, displaying the presence of significant market powers. We abduct, that the features of Tamil Nadu mobile broadband price plans, as described when introducing Insight 4 above, should be affected by some of the hierarchical structuring features, derived from the above-discussed metrics, characterising the upstream Internet market. This could be due to the pervasive presence of Autonomous Systems from larger Tier-1 and Tier-2 *Internet Service Providers*. As these *ISPs* are likely to establish more peering relationships among themselves and transit ones with their customers, the resulting market power will be reflected in higher price-costs margins and, consequently by lower affordability (see section 2.5).

Nevertheless, Giovannetti and Sigloch (2015) provide the only work that studies upstream Internet market structure while relating it to *Degree* connectivity and its resulting effects on Internet market positions in a developing country. To the best of our knowledge, no other researcher in the field of Internet Economics has merged Network Analysis and Development Studies to explore the hierarchical organisation of the upstream Internet market structure (at different analytical granularities) using active Internet periphery measurements. This reveals a relevant gap in this literature.

Moreover, the potential effect of hierarchical upstream Internet market structures on the affordability of mobile broadband for end-users in developing and emerging countries is likewise under-researched. Therefore, this dissertation focuses on methodologically extending and building on the preliminary work of Giovannetti and Sigloch (2015).

2.3.4 Graph Visualisations and Layouts

Graph visualisations

Due to its sheer size, mapping the Internet often relates to difficult abstractions of the real-world (Danesh et al., 2001). The first graph visualisation of the Internet might have been the backbone drawing of the early ARPANET from 1969 as Figure 2-3 on the next page illustrates. This graph visualisation includes the first four institutions of the ARPANET (see section 2.1.1 above) as well as their respective connections.

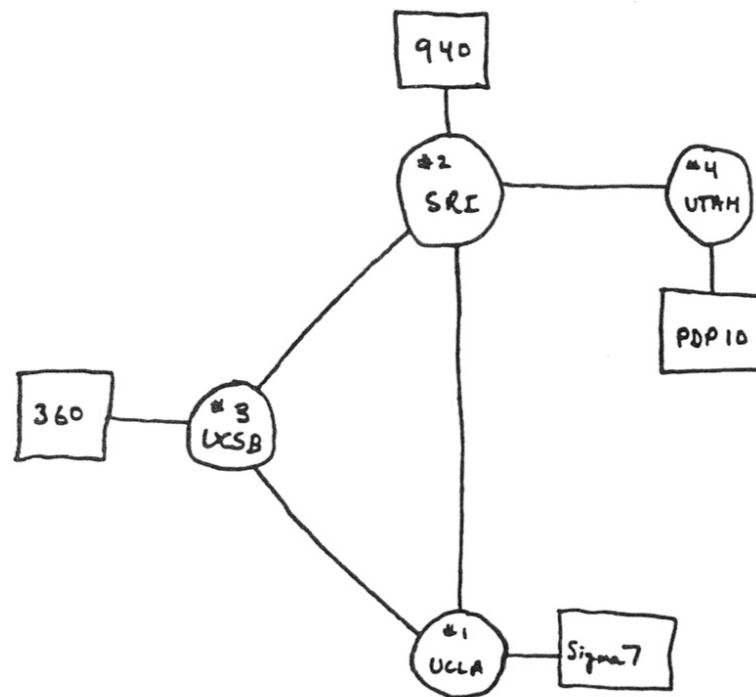


Figure 2-3: ARPANET 1969 graph visualisation, Source: The Ocp (2016).

With the increasing structural complexity of the Internet, Burch and Cheswick (1999) attempted to map the Internet by studying 88,000 Internet Protocol (IP) addresses and their associated routers and found critical indications for hop distances between their local Carnegie Mellon University and Lycos, an important search engine at that time. By using paths from a local test host containing 90,000 networks towards another host on a destination network, Cheswick, Burch and Branigan (2000) visualise network vertices using a force-directed graph visualisation layout (see below). Their visualisation reveals a number of interesting *Internet Service Providers*. However, their work also mentions the high complexity of the graph visualisation, which makes it hard for them to conclude their findings with great confidence (Cheswick, Burch and Branigan, 2000). Moreover, there is a relevant group of authors that aim to map the Internet infrastructure from a

geographical perspective. Important examples include Lakhina et al. (2003) who use the CAIDA dataset from 20 ‘Skitter’ monitors (see project description above) with the CAIDA NetGeo IxMapper. Another example is the work of Shavitt and Zilberman (2012), who utilise the DIMES database to map Point-of-Presence connectivity (vertices between networks) with the Arc GIS (2016) mapping software. Roberts et al. (2011) elaborate on one of the only approaches to map the Indian Autonomous System landscape. However, their study covers 100 countries using the CAIDA (2016b) AS-Relationship data and does not explicitly focus on the Indian landscape. Dimitropoulos et al. (2007) reveal country-level Autonomous Systems with the greatest network control. Their work presents the resulting graph visualisations in a *Circular Layout* using the Flare Toolkit (2010) for China, Russia, The Republic of Korea, The Islamic Republic of Iran, Egypt, Sweden, Ukraine, Angola and India, as well as a comparison between those graphs. Their work identifies four Autonomous Systems with a great control for the 17,98 million analysed Indian IP addresses but does not mention the names of these Autonomous Systems, while also neglecting the structural properties of their composition. However, Dimitropoulos et al. (2007) indicate that the number of Indian Autonomous Systems with great network control is fairly low compared to the other countries in their study. Their work also neglects an end-user perspective and only builds on secondary data. Notably, there are hardly any research efforts mapping upstream Autonomous System relationships and especially upstream connectivity structuring, originating from an Internet Periphery perspective. Caldarelli, Marchetti and Pietronero (2000) analyse, on the basis of the data obtained from Cheswick, Burch and Branigan (2000), some network indicators from an end-user perspective at IP granularity where they find signs of hierarchical structural ordering between end-users and providers. Tangmunarunkit et al. (2002) independently of Caldarelli, Marchetti and Pietronero (2000) or Cheswick, Burch and Branigan (2000) show that while the Internet embodies a hierarchical structuring, graphs are better modelled without explicitly constructing hierarchies. This refers to network visualisations using the Directed Acyclic Graph layout, amongst others. While Giovannetti and Sigloch (2015) explore the upstream network connectivity structure of the incumbent Bhutanese mobile broadband provider, B-Mobile, from an Internet Periphery Analysis introduced by Faggiani et al. (2012), their generated graph visualisation only covers rudimentary analysis.

Graph Layouts

Graph visualisations are considered to be well suited to display agents and their relationship information in networks (Eick, 1996). Moreover, the visualisation of network graphs helps, according to Bastian, Heymann and Jacomy (2009), to understand network structures and their data, while the process of graph visualisation analyses is best suited to follow exploratory strategies (Perer and Shneiderman, 2006). Graph visualisations are done using graph layouts that represent the spatial foundation of a visualisation, including the positioning of vertices and the edges among them. Therefore, graph layouts are used for highlighting specific but highly relevant graph characteristics (Brath and Jonker, 2015). In Figure 2-4 below we generated a random network graph visualisation of an example network consisting of 200 vertices and 1,333 edges linking those vertices using a Random Layout. The graph visualisation using the Random Layout fails, as expected given the Random Layout, to display specific network characteristics.

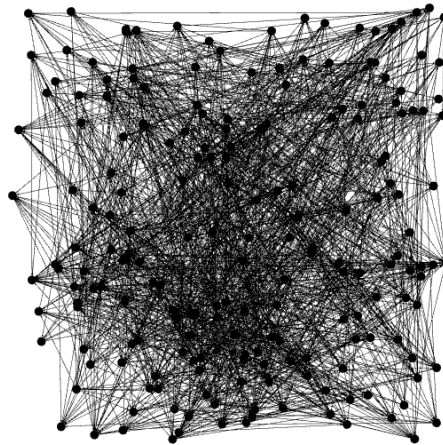


Figure 2-4: Random Layout graph visualisation with 200 vertices and 1333 edges, elaborated using Gephi (2016).

Nevertheless, the choice of a specific graph layout depends on the research questions addressed. While the literature shows a large set of possible graph visualisations, only a few are valuable for the exploratory analysis of Internet network structures. Among the visualisations considered in this work is the *Layered Layout* by Kuchar (2012), which places vertices in different layers depending on specifically chosen attributes. According

to Kuchar (2012), this layout is particularly appropriate for studying *Small-World Network* phenomena, representing (completely dense) connections between any vertices in a network. This layout, therefore, helps in testing whether or not any of our mobile broadband operators displays *Small-World Networks* characteristics. Given the importance of large Tier-1 *Internet Service Providers*, we would expect to see no *Small-World Network* effects. Nevertheless, this analysis helps to obtain indicators of the interconnection efficiency of a network. If every Autonomous System were connected to any other Autonomous System in the network, then the networks wouldn't display hierarchical structural features, which is highly unlikely considering the aforementioned tier-ordering of the Internet. Figure 2-5 below illustrates the same random example network graph visualisation as above using a *Layered Layout*.

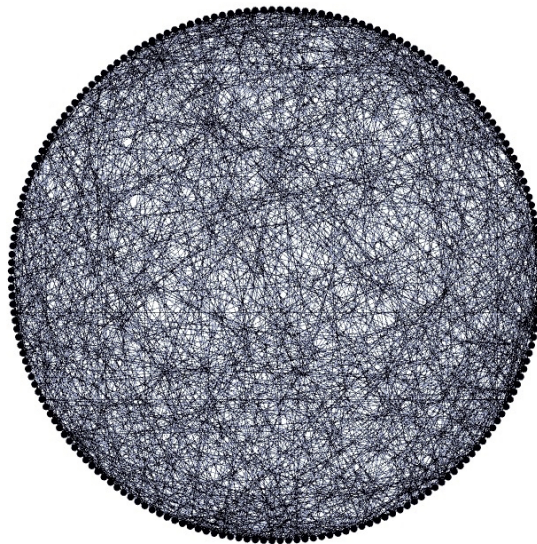


Figure 2-5: Layered Layout graph visualisation with 200 vertices and 1333 edges, elaborated using Gephi (2016).

A different graph visualisation is the *Fruchterman – Reingold Layout*, which focuses on visualisation aesthetics, meaning that edges are more or less having the same visualisation lengths while not crossing each other in the visualisation. This is arranged by applying forces to the edges and vertices based on their relative position in the network (Fruchterman and Reingold, 1991). These forces are applied using spring-like attractions using the *Hooke's law* of Physics. This graph layout, therefore, helps when analysing the importance of specific edges in a network. Chan et al. (2003) use the *Fruchterman-*

Reingold, rather than other layouts, to visualise the structure of a *Border Gateway Protocol* routing networks, while proposing a new layout. Moreover, they consider the *Fruchterman-Reingold* Layout as particularly useful to capture and visualise the presence of *power-law degree distributions* in networks. Figure 2-6 below depicts the same example network graph visualisation as above using a *Fruchterman-Reingold* Layout.

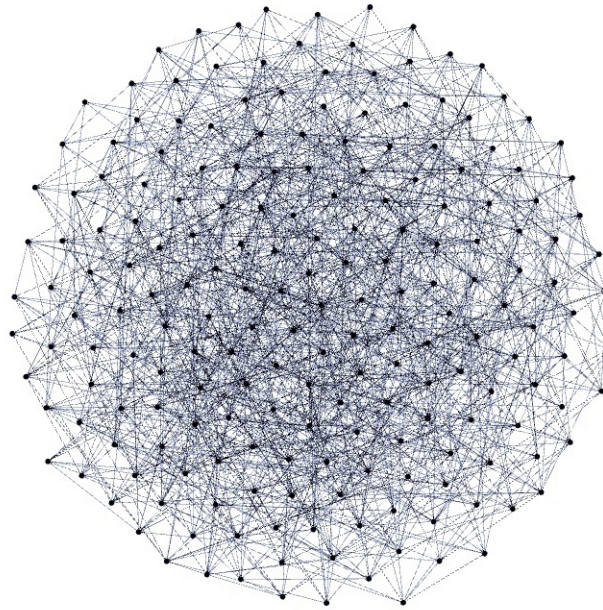


Figure 2-6: Fruchterman-Reingold Layout graph visualisation with 200 vertices and 1333 edges, elaborated using Gephi (2016).

To increase the intuitive usage of general layouts, Jacomy et al. (2014) introduce the *Force Atlas 2* layout. This layout is considered to be useful in helping an intuitive spatial visualisation of networks. Compared to the *Fruchterman Reingold* layout, the *Force Atlas 2* layout shows better performance and usability with strongly clustered networks. This is important since performance ultimately adds to the readability of the graph visualisation. Moreover, the *Force Atlas 2* layout employs avoidances of vertex overlap, which is particularly interesting when trying to identify vertex clusters or white spaces of unconnected vertices in the network structure. In terms of its application, Hasani and Mehdipour (2015) use the *Force Atlas 2* layout for visualising traffic in an Internet Protocol (IP) address network. Figure 2-7 below illustrates the same random example network graph visualisation as above (200 vertices and 1333 edges) using a *Force Atlas 2* Layout.

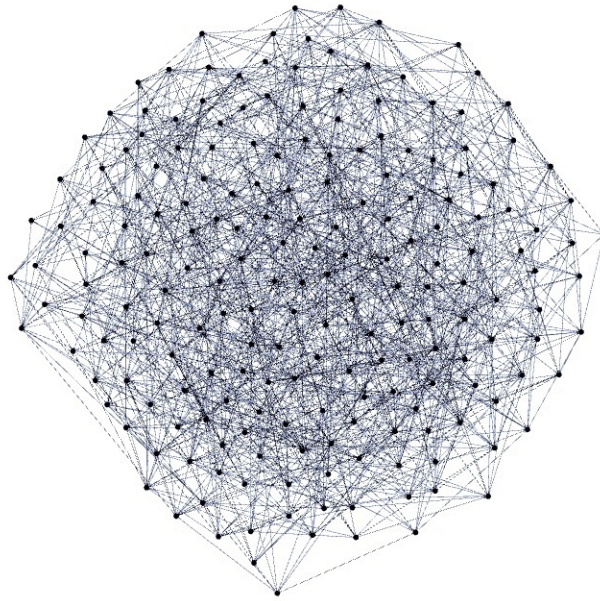


Figure 2-7: Force Atlas 2 Layout graph visualisation with 200 vertices and 1333 edges, elaborated using Gephi (2016).

More closely related to the analysis of Autonomous System networks, Alvarez-Hamelin et al. (2005b) introduce and use a *k-core decomposition* for the World Wide Web and Internet analysis. Carmi et al. (2005) and Alvarez-Hamelin et al. (2008) use the *k-core decomposition* in communication networks such as the Internet at Autonomous System granularity. The *k-core decomposition* separates the network vertices into so-called *k-cores* (see coloured *k-cores* in Figure 2-8 below), or sub-graphs, based on the given connection densities amongst vertices. This means that the most densely connected vertices would be situated in the highest *k-core* of the network visualisation, whereas less dense connected vertices are situated increasingly in the periphery of the visualisation. Hence, the *k-core decomposition* indicates the most important hierarchical vertices of a given network. Figure 2-8 below depicts the same random example network graph visualisation as above (200 vertices and 1333 edges) using the *k-core decomposition*.

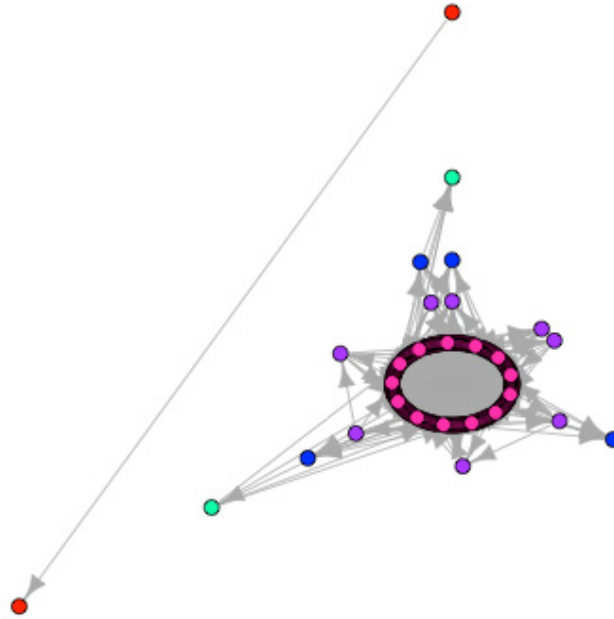


Figure 2-8: *k*-core decomposition graph visualisation with 200 vertices and 1333 edges, elaborated using R (2016).

Insight 11: Given its distinct applicability to study the structure of the Internet (see Alvarez-Hamelin et al., 2005b), we consider the *k*-core decomposition and its resulting graph visualisation as the best algorithm to discover influential Autonomous System vertices in our networks. Given the economic nature, we expect that the most densely connected Autonomous Systems being Tier-1 *Internet Service Providers*. Moreover, we expect that other graph layout visualisations, such as Force Atlas 2, provide valuable exploratory insights on structural features. These indications will be useful to explain and compare the three mobile broadband operator graph visualisations and their general structural features. Our work is the first to apply such a broad spectrum to the exploratory analysis of active Internet periphery data.

2.4 Key Gaps in the Literature

Based on the Literature Review above, our dissertation embarks on a journey to explore two identified key gaps in the literature. To date, and to the best of our knowledge, no research studied in-depth the following two themes, while linking them to the necessary research disciplines:

- The study of hierarchical structuring of mobile broadband operator networks and the resulting structural *Internet Service Provider* bottlenecks in the upstream Internet market in India (or the state of Tamil Nadu) using Network Analytical methods upon data obtained from active Internet periphery measurements.
- The econometric analysis of the potential effect of those hierarchical upstream Internet market structures on the affordability of mobile broadband for end-users in developing and emerging countries.

These key gaps are exploratory and will be studied, in the remaining chapters, making use of *Working Hypotheses* being derived from the insights gathered from an abductive approach to research (see section 3.2).

2.5 Abduction

Abductive inferences are logical inferences for the purpose of finding the most likely explanations of surprising facts. The starting point of an abductive inference is provided by observational insights that flashed like ‘mental heat-lightning’ (Peirce, 1974). We extracted 11 insights highlighted throughout the Literature Review above. These insights conveyed valuable methodological information to be used in our iterative research process (see section 3.2). Moreover, these insights informed our abductive inferences.

Hence, we start our inquiry process with the surprising key insights that we extracted from the Literature Review above. The two insights extracted from the Literature Review informed our thought process on the abductive inference. Therefore, we abduct that:

- If the Bhutanese mobile broadband operator displays features of a hierarchical upstream Internet market structure, the Tamil Nadu mobile broadband operators may also display features of a hierarchical upstream Internet market structure.
- These hierarchical upstream Internet market features may have a relation to the indicated low affordability and Quality of Service of mobile broadband in Tamil Nadu.

Next, we specify the following *Working Hypotheses* to explore and pragmatically understand our abductive inferences. The specified *Working Hypotheses* are then analysed in the upcoming analytical steps of our dissertation (Chapter 4 and Chapter 5).

WH1: ‘The Tamil Nadu mobile broadband operators’ upstream Internet market structure displays features of a hierarchical ordering’.

WH1.1: ‘The Tamil Nadu mobile broadband operators rely on an identified set of specific Internet Service Providers for their upstream connectivity’.

WH1.2: ‘Studying the Tamil Nadu mobile broadband operators from an Internet-Periphery perspective indicates previously hidden upstream AS relationships’.

WH2: ‘Tamil Nadu mobile broadband operators that show signs of a hierarchical upstream Internet market structure offer less affordable mobile broadband price plans to an end-user’.

WH3: ‘Those Tamil Nadu mobile broadband operators that show signs of a hierarchical upstream Internet market structure provide a lower quality of service to an end-user’.

2.6 Conceptual Framework

The following conceptual framework organises the 11 identified key insights derived from the literature review and the key gaps in this literature. Moreover, the conceptual framework, represented in Figure 2-9 below, relates and provides the relevant conceptual structure for exploring the identified Working Hypotheses derived in section 2.5 (black arrows). Figure 2-9 below, hence, provides both a valuable overview of this dissertation and a mapping of how our anticipated findings relate to each other (green arrows).

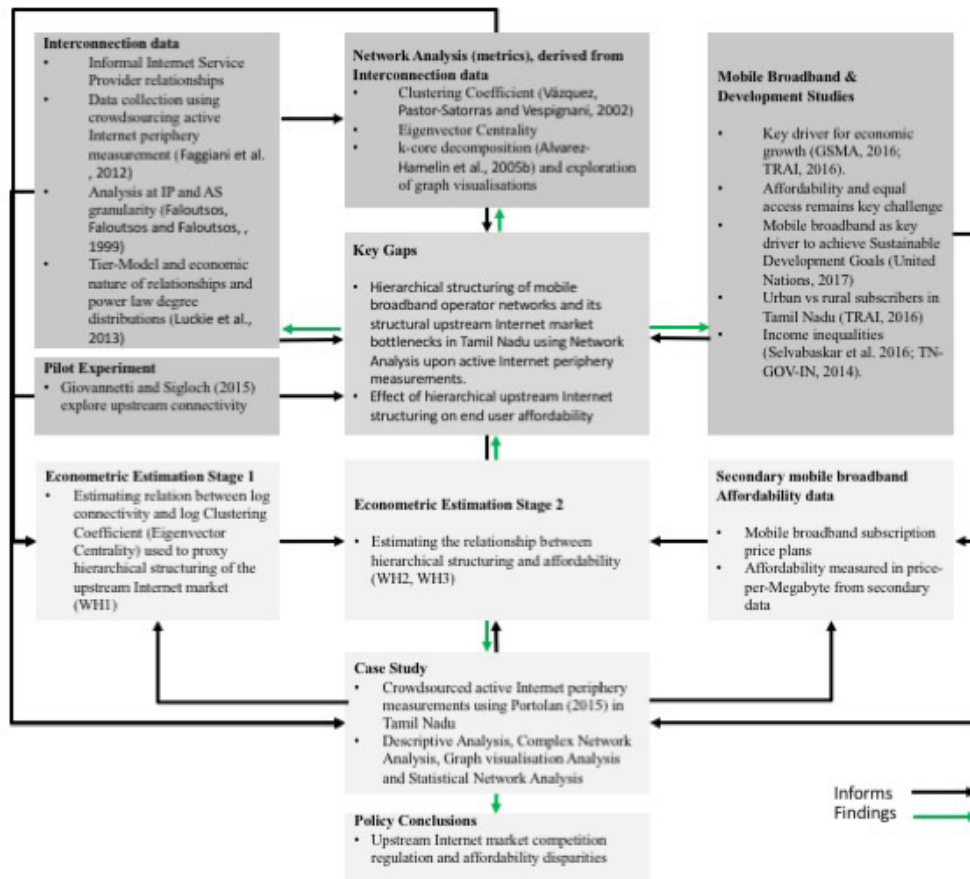


Figure 2-9: Conceptual Framework.

2.7 Research Aims and Objectives

The following Research Aims and Objectives are derived from our abductive inferences and the specified *Working Hypotheses*, discussed in section 2.5 above. Again, these *Working Hypotheses* were extracted from insights emerging from the key gaps in the Literature Review rooted in the literature of our three relevant disciplines: Network Analysis, Development Studies and Computer Sciences, as depicted in the Conceptual Framework in Figure 2-9. Here, we first state our Research Aims in section 2.7.1 below before listing our Research Objectives in section 2.7.2.

2.7.1 Research Aims

Emerging from the identified gaps in the key literature, see section 2.4, our two main research aims are:

- To discover the hidden hierarchical structuring of the upstream Internet market for the Tamil Nadu mobile broadband operators using active Internet periphery

measurements.

- To reveal that this structuring is a key determinant for Tamil Nadu's mobile broadband affordability and Quality of Service for end-users.

The first one of these research aims is expected to shed light on the structuring of the upstream Internet market, a feature that is greatly hidden for conventional analysis methods but clearly visible using active Internet periphery measurements using smartphones, as Giovannetti and Sigloch (2015) showed. The second one of these research aims is grounded on the belief that the economic nature in the Tier-organised upstream Internet market is intrinsically asymmetric, meaning that large Internet Service Providers display stronger bargaining powers towards their downstream partners.

2.7.2 Research Objectives

Given our abducted *Working Hypotheses* and the above Research Aims, we propose the following set of research objectives. Each objective serves as a milestone for gaining the necessary information to assess our *Working Hypotheses*.

Objective 1: To collect *traceroute*-based upstream connectivity data for the Tamil Nadu mobile broadband operators using the active Internet periphery measurement tool Portolan (2015). This shall uncover hidden features of the upstream Internet market as such features are usually not visible using measurements from the Internet core (see literature discussed in section 2.3.1).

Objective 2: To describe, explore, analyse and understand the collected *traceroute* upstream connectivity data at Internet Protocol and Autonomous System granularity. This follows best practices on data collection granularities as discussed in the literature review (see section 2.3.2) and in particular by Faloutsos, Faloutsos and Faloutsos, (1999).

Objective 3: To prove the existence of *power-law degree distributions* at Internet Protocol and Autonomous System granularity for our collected data in the Tamil Nadu case studies. These distributions would be a clear indicator for hierarchical structuring in the upstream Internet market as Faloutsos, Faloutsos and Faloutsos, (1999) and Dall'Asta et al. (2005) show. The identification follows best practices employed in the Computer Sciences realm (see section 2.3.2).

Objective 4: To generate, explore, analyse and compare the hierarchical features of

mobile broadband operator networks using the most relevant *Complex Network Analysis* metrics at Internet Protocol and Autonomous System granularity. In detail, our objectives are to

- (i) identify markers of hierarchical upstream Internet market structure and
- (ii) uncover the most relevant business relationships between *Internet Service Providers* by fusing the generated networks data with the additional secondary CAIDA (2016b) Autonomous System relationship data.

This objective relates to the methodological approach used in Giovannetti and Sigloch (2015). Using new markers for the identification of hierarchical upstream Internet market structures, namely the Eigenvector Centrality, our objective is to additionally relate to the work of Vázquez, Pastor-Satorras and Vespignani (2002), who neglected this centrality metrics while focussing on the Clustering Coefficient.

Objective 5: To generate graph visualisations based on the collected *traceroute* upstream connectivity data using the Open Source Network Exploration Tool *Gephi* (2016) and the Statistical Computing Software *R* (2016). These visualisations will help to identify the underlying structures of the present upstream Internet markets. Using these analysis tools, we add to the work of graph visualisation analysis using algorithms such as Barabási and Albert (2002) as discussed in section 2.3.4.

Objective 6: To compare the identified upstream Internet market properties and the identified business relationships of the Tamil Nadu mobile broadband operators. This helps us to identify the hidden features of the upstream Internet markets, representing novel insights on their business relationships, a topic analysed in section 4.4.

Objective 7: To further explore and display the identified upstream connectivity properties and business relationships by using *Graph Visualisation Analysis* techniques, as discussed in the literature reviewed in section 2.3.4.

Objective 8: To develop a set of econometric models for testing our *Working Hypotheses* about the insights and obtained evidence and to explore the effects of the upstream Internet market structure on the affordability of Tamil Nadu price plans, as discussed in the vast body of multidisciplinary literature reviewed above.

Objective 9: To propose additional hypotheses to further test our findings for causality

through future, explanatory research. This closes the loop of our abductive approach to research. The new hypotheses generated are stated in section 7.1.

3 METHODOLOGY

‘To a pragmatist, the mandate of science is not to find truth or reality, the existence of which are perpetually in dispute, but to facilitate human problem-solving’, (Powell, 2001, p.884).

In this Chapter, we aim to set out the philosophical assumptions that reinforce the nature, scope and limitations of this dissertation. Based on the *Working Hypotheses* from the Literature Review (see section 2.5), we start by stating the pragmatist epistemological and ontological stance assumed towards reality and the nature of knowledge. Next, follows an explanation of this dissertation’s abductive approach to research, which ultimately informed our choice of the case study research design and the multimethod research. This section then covers details on the selection of time horizon and the crowdsourced collection of cross-sectional upstream *traceroute* data and follows up with its necessary materials and preparation tasks. Finally, we discuss the different *Complex, Graph Visualisation* and *Statistical Network Analysis* that were used to explore the set of *Working Hypotheses* and end by stating considerations on Ethics, Biases, Reliability, Validity and Generalisability.

3.1 Pragmatism

In the following, we discuss our epistemological and ontological choices, before defining the research design, adopted to inquire this dissertation’s research problem.

Paradigms are traditionally shared beliefs within certain communities of researchers such as post-positivists or constructivists. As a pragmatist, my focus lies on the characteristics of fruitful approaches to inquiry. Here, Pragmatism is a radical departure from traditional philosophical arguments about the nature of reality and the possibility to experience truth (Morgan, 2014). Hall (2013, p. 19) finds that pragmatism offers ‘an alternative epistemological paradigm’, a new worldview where knowledge consists of assertions as results from actions and the experience of outcomes (see also Dewey, 1941).

The following discussion sheds light on the relevance of the Pragmatism paradigm to study the economic effects of hierarchical upstream Internet market structures, from both a Social Science and Computer Science induced perspective.

'Fortunately, nobody owns the Internet, there is no centralized control, and nobody can turn it off. Its evolution depends on rough consensus about technical proposals, and on running code.' (Carpenter, 1996).

The Internet is considered to be a system, or network, of interconnected networks, consisting of 'bare-metal' machines whose ports are linked between each other using physical transmission links and the software-based Internet Protocol (IP) Suite. The structure of the Internet can be studied at different levels of abstraction or granularities (see section 2.3.2). Unlike some researchers, we set the granularities for this dissertation at Internet Protocol granularity (Machine A talks to Machine B) and the Autonomous System granularity (Network Organisation A talks to Network Organisation B). The connections between machines (mostly routers), or networks of machines, are usually studied and visualised through different types of Network Analysis where networks are visualised in graphs. Therefore, each of these network graph visualisations represents a modelled copy of the real world. Such modelled realities are usually incomplete as different data and methodologies are yielding divergent views on the Internet (Mahadevan et al., 2006). An in-depth analysis of these modelled realities involves a lot of detailed inquiry since visualising and analysing these networks broadens a researcher's awareness towards the entities involved, their relationships and their respective roles under given circumstances. This is also a valid notion in the Pragmatism paradigm. By linking abduction to computation and philosophy, Josephsen and Josephsen (1996) compare the approach of abductive reasoning to 'detective work', where researchers collect related 'facts' about entities in some given circumstances. Therefore, Pragmatism seems especially suitable for an exploratory analysis of the hierarchical upstream Internet market structure, where we explore *traceroute* data under the light of abductive reasoning until a plausible story for the data and hence an econometric model suitable for the given reality emerges.

According to Powell (2001), a true proposition facilitates paths of discovery to come about a realisation of 'pragmatic truth', or follows the scientific discovery as introduced by Peirce (1878). By following a pragmatist approach to research, our ontological considerations refer to the ways in which we may or may not justify what we assert about the ontological consideration being that 'truth' may not be all-embracing. Cotton, Tashakkori and Teddlie (1999) argue that pragmatism creates the need to triangulate findings. We distanced our research from this argument by saying that our choice for

pragmatist approach to research, following an exploratory research design, aimed to generate feasible hypotheses in a case study setting. Therefore, these hypotheses ought to be re-tested for causality in future explanatory research. Hence, triangulation is not considered an issue in this research since the construct of real truth was not to be argued here. However, the creation of knowledge induced an active process of inquiry, which was created using a continual back-and-forth movement between our beliefs and corresponding actions.

3.2 Abductive Approach to Research

The chosen research approach aimed to determine a systematic thread for quantitatively exploring the identified research problems. According to Peirce (1878), abductive approaches to research are divided into the following three stages:

1. Logical Inference (abduction) that ought to be explained as a search for a meaningful rule.
2. Definition of plausible Working Hypotheses (see section 2.5).
3. Testing and hence either pragmatically verifying or falsifying the defined Working Hypotheses by means and choice of research methods.

Both the logical inferences and our abducted *Working Hypotheses* are stated in the Literature Review. The testing and pragmatic verification or falsification follows in the upcoming Chapter 4 and Chapter 5. In theory, to establish ‘all-embracing truth’, one may repeat these exploratory stages ad infinitum (Dewey, 1941). However, to come about ‘pragmatic truth’, the *Working Hypotheses* of our research are gradually expanded throughout Chapter 4 and then statistically tested in Chapter 5. Linked with our ontological considerations, it was not possible to achieve certainty as to our abducted *Working Hypotheses*’ validity. Nevertheless, a strategic choice upon a single case study strategy allowed us to gain pragmatic validity. By an interpretation of collected data, abduction consists of assembling and discovering features. Abduction is a result of an intellectual process that flashes like ‘mental heat-lightning’, where rule hampers the thought process (Peirce, 1974). Therefore, in following this mental effort, one first has to discover or invent a process of reflection, while utilising general thinking patterns as Figure 3-3 in section 3.4 illustrates.

3.3 Exploratory Research based on a Single Case Study Strategy

Exploration seemed to be the most plausible research design when searching for a meaningful rule. This section first covers our reasoning for choosing an exploratory-quantitative research design. Secondly, it discusses the suitability of a single hypothesis-generating case study strategy.

3.3.1 Exploratory-Quantitative Research Design

Research always begins with curiosity, representing a seed of creativity. Following this curiosity, most research types in the Social Sciences face steps of best-guessing inquiry, stumbling around information, searching and examining gaps in knowledge, or investigating hunches of rather loosely found insights to categorise and report what one has learned. While all of these steps can be related to exploratory research, Stebbins (2001, p.vii) notes that exploration is falsely regarded as an ‘outmoded process’ and suggests each research in the Social Sciences is somewhat exploratory. Exploratory research is highly applicable where the field of interest shows a lack of preliminary research and the research problems cannot be clearly defined (Stebbins, 2001), while the nature of exploratory research is informed by theory, rather than driven by theory (Waters, 2007). Moreover, the exploration of data, the findings of patterns, and the suggestion of hypotheses relates to the nature of knowledge and reality in the Peircean logical system of Pragmatism (Yu, 2006) whereas quantitative researchers employing exploratory processes relate to exploratory-quantitative research (Stebbins, 2001, p. 8). The ultimate objective of exploratory research is the investigation of key issues and key variables as distinct phenomena for the purpose of suggesting hypotheses that can be feasibly tested for causality by following explanatory research (Streb, 2016).

Hence, as a pragmatist, our choice of a quantitative exploratory research design was grounded in the following reasoning. Two of our three literature disciplines showed a clear lack of detailed preliminary research and therefore valuable gaps in existing knowledge. The concepts that applied to our research disciplines seemed clear from one level of abstraction but rather unclear from another level of abstraction (Internet structure granularities as stated in section 2.3.2). Furthermore, none of our *Working Hypotheses* were sufficiently covered in preliminary research in all of the three literature disciplines, but could be studied in a natural, real-world case study setting. Therefore, it was assumed that an exploratory research design based on a hypothesis-generating case study was the most appropriate research design for seeking pragmatic answers for both the

understanding of our logically inferred *Working Hypotheses* and the solving of the research problem. In doing so, we aimed to generate results that ought to be further tested for causality in future explanatory studies.

Following the nature of quantitative exploratory research, this Research Design should bring about well-informed insights. According to Peirce (1878), knowledge is fallible in nature. However, pragmatists are satisfied with stable beliefs. By exploring our research topic at varying levels of depth, the exploratory research under the light of pragmatism helped us to explore the *Working Hypotheses* at hand (section 2.5). Suitable for our pragmatist nature and the rather low coverage of our research problem in the literature, Brown and Saunders (2006, p.43) note that exploratory research ‘tends to tackle new problems on which little or no previous research has been done’. Saunders et al. (2009) further state that exploratory researchers need to be willing to change their directions based on the occurrence of new data or insight. Nargundkar (2003, p.41) mentions that exploratory researchers may, therefore, work as ‘methodologically’ as possible. This issue was taken into further consideration when choosing the most suitable research methods for this case study. Hence, our stated research methods were a result of repeated changes of directions in this thought process. Notably, exploratory research ‘simply’ explores the stated research questions. In considering the differences to conclusive studies, Sandhusen (2001) says that exploratory research will result in a range of causes and alternative options to find solutions to a particular problem, whereas conclusive research further aims to identify the final causality to an existing problem. However, finding causality is considered pragmatic, while not referring to it as final or generalisable causality. Hence, exploratory research forms the basis of more conclusive research and might even help to determine suitable research designs, sampling methodologies, or data collection techniques (Singh, 2007, p.64).

By following an abductive approach to research, our exploratory research design commenced with the abductive inferences of key issues that emerged from the three reviewed literature disciplines (section 2.4). This was followed by a definition of research aims and objectives, which included these key issues that ought to be explored (section 2.6). On that basis, the next step was to choose the most appropriate research strategy for exploring the *Working Hypotheses*. The strategic choice fell to a hypotheses-generating single case study strategy as the next section introduces.

3.3.2 Hypotheses-Generating Single Case Study Strategy

For the purpose of this dissertation, we analysed thirteen key characteristics stated by Stake (1995), Yin (2003), Flyvbjerg (2006) and Bryman (2012), that ought to be given when choosing a case study strategy (see Table 9-5 in Appendices). According to this analysis, case study methodologies are useful when the subjects of interest can be analysed in a natural setting, a strong theoretical base does not support the research phenomena and the phase of research covers hypothesis generation by utilising exploratory research designs (Yin, 2003). Based on this dissertation's Research aims and objectives (section 2.6), the subject of inquiry was the focus of interest while Exploratory-Quantitative Research Methods were used. The phenomena to be studied were best explored in a natural setting while the upstream *traceroute* data was collected using a cross-sectional crowdsourced data collection method from an Internet periphery or end-user perspective (see section 3.3.4 below). According to Flyvbjerg (2006), typical case studies are best when the objective relates to achieving a considerable amount of information on a given problem or phenomenon. Our case study strategy examined different subjects that were studied intensively from varying perspectives, while also fusing the collected *traceroute* dataset with various secondary data sets (in Chapter 4 and 5). As pragmatists, we had a healthy attitude towards exploring the given phenomena. Neither dependent, nor independent variables (covariates) were specified upfront but were results of the inquiry process. Moreover, we used our interpretive and integrative abilities when reporting evidence. Changes in the case study strategy and the inquiry process occurred naturally and led to deeper and more valuable insights of the exploration.

Single Case Study Selection

Case studies may cover an

'...analysis of ... institutions, or other systems that are studied holistically by one or more methods. The case that is the 'subject' of the inquiry will be an instance of a class of phenomena that provides an analytical frame – an object – within which the study is conducted and which the case illuminates and explicates.' (Thomas, 2010).

Following this definition, we set the strategic choice of a case study description to bring about insights for exploring our abducted *Working Hypotheses* as follows: The case, or object of interest, were the structural properties of the upstream Internet market structures,

originating from three distinct Tamil Nadu mobile broadband operator SIM cards (Aircel, Bharti Airtel and Vodafone, see section 3.3.4 below). This interest was strongly influenced by our observed insights in the Literature Review, the abducted *Working Hypotheses* (see section 2.5), our research aims and objectives (see section 2.6), and the evidence from the pilot experiment by Giovannetti and Sigloch (2015) for the incumbent mobile broadband provider, B-Mobile, in Bhutan.

Moreover, our case study was informed by the theory for structural properties of the Internet, but not controlled by it. This theory provided a valuable analytical frameset to explore our three mobile broadband operator subjects of interest, which set the lens that limited our view on both knowledge generation and problem solving. The upstream Internet access market structure may be of vital importance for understanding mobile broadband affordability and the Quality of Service of the access from the Internet periphery. Moreover, these structural features may be of special relevance for the state of Tamil Nadu in the low-middle income country, India, which lags behind regarding Internet access and also faces great urban-rural per capita income and gender disparities. Therefore, the strategic choice to explore three Tamil Nadu mobile broadband operators allowed for a valuable assessment of these hidden structural features of the upstream Internet market. Hence, this case study followed a multimethod research design for assessing the described structural features in the most conducive way. The following section elaborates on the study design for our selected single case study.

3.3.3 Cross-Sectional Study Design

Time is an important element in research (Trochim, 2006). A cross-sectional design refers to both qualitative and quantitative research where phenomena of many subjects are studied at a specific point in time and in great detail. We chose a cross-sectional study design for exploring the upstream Internet market structure of the three mobile broadband operators in Tamil Nadu. Hence, our collected data referred to network connectivity information from Aircel, Bharti Airtel and Vodafone at one point in time using a single measurement campaign. This measurement campaign was chosen to cover a relatively short data collection period to avoid computational problems of large datasets at a later stage, while still collecting enough *traceroute* data to elaborate a pragmatic and meaningful insight. The following sections describe the data collection technique while further elaborating on the necessary materials and equipment, the data collection preparation and the specific timing of the measurement campaign.

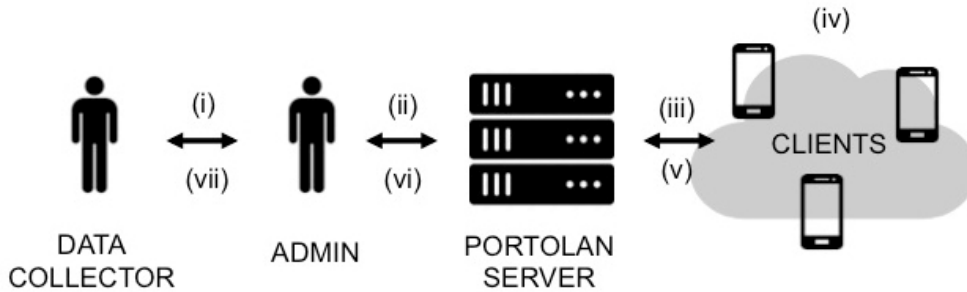
3.3.4 Mobile Crowdsourced Primary Data Collection

Crowdsourcing refers to the strategy to solve large-scale problems by utilising existing resources from the masses. Crowdsourcing approaches are considered to be a very effective when a solution relies on performing tasks on a larger scale (Faggiani et al., 2013). There are currently three notable mobile active Internet periphery measurement applications to capture Internet structural data from an Internet periphery perspective (see section 2.3.1):

- NetRadar (2015), focusing on the mobile broadband operator coverage and the comparison of accessing devices.
- OpenSignal (2015), which aims to capture signal strengths of mobile broadband operators.
- Portolan (2015), which aims to discover the topology and structure of the Internet through utilising *Paris traceroutes*.

While traditional Internet Topology mapping efforts often rely on a top-down and passive data collection approach, the data collection using the Portolan (2015) provides Internet measurements using a unique bottom-up, or active Internet periphery (end-user) perspective (Faggiani et al., 2012). We chose the Portolan (2015) application over its competitors', due to its focus on capturing *traceroutes*, allowing us to measure the upstream Internet market structure through utilising Network Analytical Methods, which represents the main interest of this dissertation (Portolan (2015) selects, for every traceroute, randomly-chosen destinations). Moreover, the applicability is already tested in a preliminary pilot experiment by Giovannetti and Sigloch (2015). The necessary preparatory steps taken for collecting the *traceroute* data using Portolan (2015) are described in the section after the next below. Figure 3-1 below illustrates the flow of the data collection in detail. Here, the data collector (researcher) first arranged the data collection campaign with the Portolan (2015) Network Tools Administrator (step (i) in Figure 3-1). This is followed by an event storing of the data collection campaign in the software of the Portolan (2015) server by the Network Tools Administrator. The server orchestrated the measurement campaigns (Faggiani et al., 2012), assigning so-called measurement campaigns to the specified Android smartphones that automatically collected the *traceroute* data on the set campaign dates. The *traceroute* data was then automatically collected by the Portolan (2015) Server, fused together and stored on a database where the respective Network Tools Administrator was able to obtain the

collected *traceroute* data (step (vi) in Figure 3-1). Next the collected *traceroute* data were sent to the data collector by the Network Tools Administrator (step (vii) in Figure 3-1), including information on how to classify the obtained files (see section 4.1, Figure 4-1).



Key

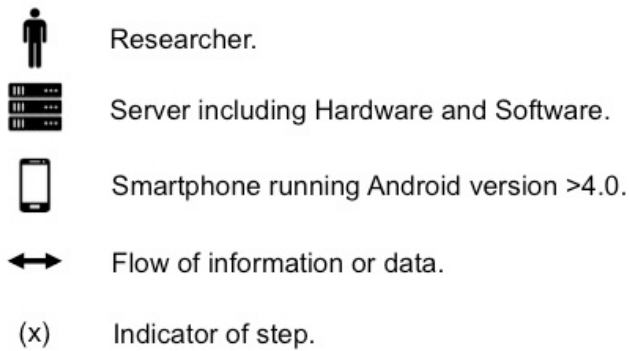


Figure 3-1: Overview of data collection.

Materials and Equipment

For the purpose of collecting the upstream connectivity *traceroute* raw data using a mobile crowdsourced data collection approach, the data collector (researcher) had to organise three Android smartphones since the Portolan (2015) application was (at the time) solely available for devices operating the Google Android Software > Version 4.0. These smartphones were of varying prices, and with decreasing cost order the brands were Sony, Micromax, Lava and Karbon. Unfortunately, only the former two were able to maintain a stable configuration of Portolan (2015). Furthermore, to collect the *traceroute* data for the purpose of exploring the stated *Working Hypotheses*, the data collector (researcher) had to organise Tamil Nadu mobile broadband operator SIM cards. As mentioned in the Literature Review above (section 2.2.2), the Mobile Service Area of Tamil Nadu is separated into two Mobile Service Areas covered by four Indian mobile

broadband operators (Aircel, Bharti Airtel, BSNL, and Vodafone), three of which the data collector was able to obtain local SIM cards for, namely Aircel, Bharti Airtel, and Vodafone. A SIM-card for the fourth mobile broadband operator BSNL (Bharat Sanchar Nigam Ltd.) could not be obtained due to local regulations for the issuing of SIM cards to locals and foreigners. This choice of materials and equipment may result in a selection bias and has therefore been taken into consideration when reporting the results (see section 3.6.1 below). Some of the chosen low-end smartphones in particular did not seem to properly collect the *traceroute* observations at times, which potentially indicates real-world connectivity situations.

Data collection preparations

Once the Android smartphones and SIM cards of the local Tamil Nadu mobile broadband operators were obtained, the next preparation tasks for the researcher were:

- To collaboratively organise the study plan for 01st March – 05th March 2015 with the Network Tools Administrator at the Instituto di Informatica e Telematica (IIT) at the University of Pisa in Italy by email. This organisation included the setting of the measurement campaign given the anticipated planning. Moreover, the organisation included the transmission of necessary information for the Portolan Server to trace the correct smartphones. Lastly, we organised the data transmission of the collected *traceroute* data from the Portolan Server from the Network Tools Administrator to the data collector.
- To organise travel and accommodation from Cambridge, UK to the Indian Institute of Technology Madras (IITM) campus in Chennai, Tamil Nadu, India.
- To collect the respective Android smartphones and SIM cards before being on-site in Chennai, Tamil Nadu.
- To organise a local driver for the purpose of travel assistance for each day of the chosen data collection period.
- To download and install the Portolan (2015) from the Google Play Store.
- To prepare the settings of the Portolan (2015) Network Sensing Architecture Android application for our chosen *traceroute* data collection purposes.

Location and Data collection times

The data collection took place in the period of 01st March 2015 – 05th March 2015, while travelling between the Indian Institute of Technology Madras campus in the urban area

of Chennai, through the rural areas of Tamil Nadu, 45 miles to the distant historical city of Kanchipuram (see Figure 3-2). The researcher, hence, collected *traceroutes* from both urban and more rural areas in Tamil Nadu, India. Before each of the planned daily commutes by car and foot, the researcher made sure that the batteries of the three Android smartphones were charged during night-time hours. This represented a normal end-user behaviour and prevented unforeseen smartphone shutdowns due to flat batteries. During the data collection commutes, it was important to mimic an end-user's usual smartphone usage behaviour. Hence, our case study covered the following use-cases of locals or tourists that are commuting or travelling to Chennai, the urban outskirts or the city of Kanchipuram (for religious events or ritual traditions such as the holy pilgrimage).

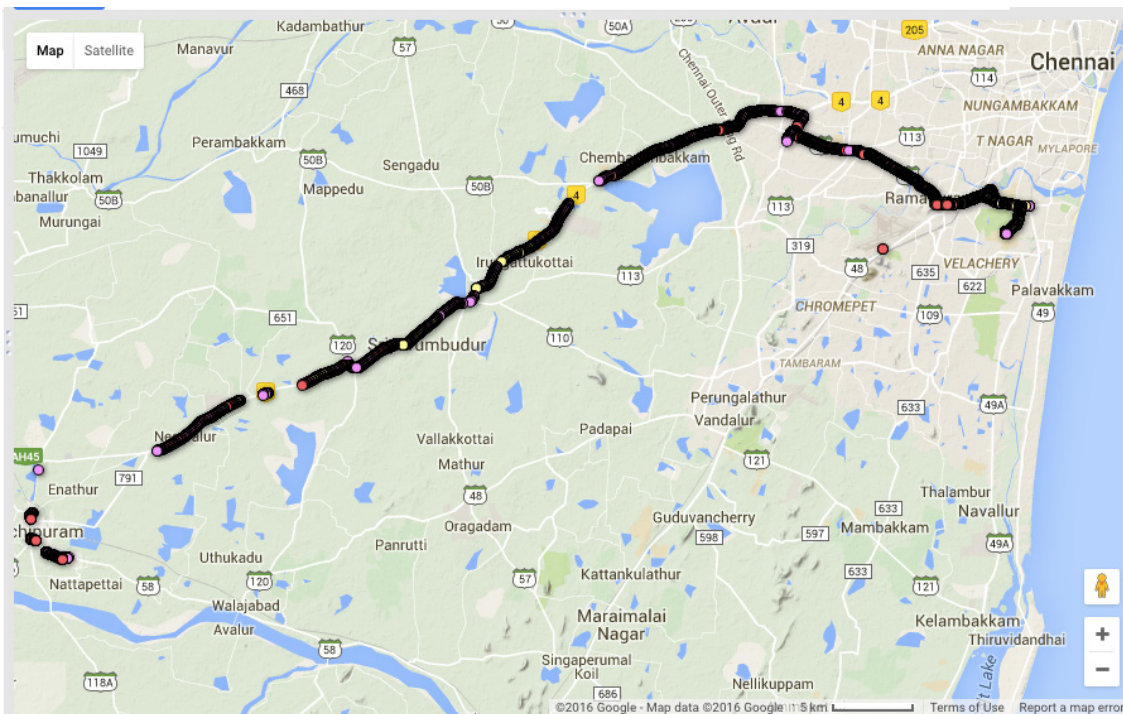


Figure 3-2: Traceroute hop observations as obtained through Portolan (2015) plotted on a Google Maps.

While carrying the smartphones during the travel commutes, the Portolan (2015) application automatically collected 57,122 unique *traceroute* observations. These *traceroute* observations contained a total number of 731,200 Internet Protocol (IP) address hop observations since each *traceroute* observation contains a multitude of Internet Protocol (IP) addresses that are traversed from any connection measurement source to a random-assigned destination. Table 3-1 below lists the distribution of collected *traceroute* hop observations per data collection day.

Observations per data collection day		
Date of data collection (YYYY, MM, DD)	Number of collected <i>traceroute</i> IP address hop observations	In percentage of total collected IP address hop observations
2015-03-01	236,805	32.39%
2015-03-02	113,431	15.51%
2015-03-03	119,373	16.33%
2015-03-04	134,621	18.41%
2015-03-05	126,970	17.36%
Total	731,200	100%
Key IP: Internet Protocol. YYYY, MM, DD: Year, Month, Day.		

Table 3-1: Observations per data collection day.

3.4 Network Analytical Multimethod Research

This section introduces the Network Analytical multimethod research that were used to analyse the collected 57,122 unique *traceroute* observations, containing a total number of 731,200 IP address hop observations, as described above.

Given the adopted pragmatic approach, we granted ourselves the freedom to choose any exploratory-quantitative research method that fruitfully helped to explore the set *Working Hypotheses*. In following Dewey's model of inquiry, or process for the production of knowledge, our aim was to generate 'warranted assertions', where warrants are outcomes of inquiry, or outcomes of using our belief in practice (Dewey, 1941). Thus, we were not separating knowledge from our practice of doing (e.g. analysing). Figure 3-3 below illustrates this process.

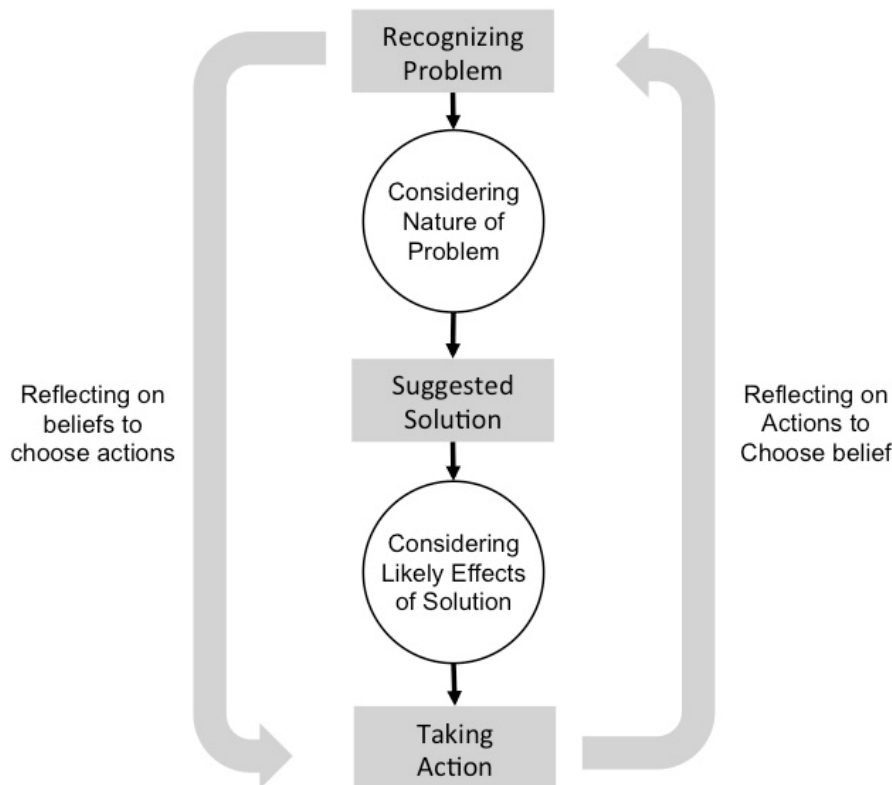


Figure 3-3: Warranted assertions process to inquiry, Source: Dewey (1941).

Our four utilised research methods, namely *Descriptive Statistics*, *Complex Network Analysis*, *Graph Visualisation Analysis* and *Statistical Network Analysis* (further discussed in this and the following sections) were a result of following Dewey's model of inquiry repeatedly until we found conducive methods to obtain a clearer picture for understanding and exploring the abducted *Working Hypotheses*. This means that every choice of subsequent research method was informed by reflecting on our undertaken actions to recognise questions, patterns and problems.

In the following sections, we discuss the choice of using the defined research methods. First, we present the *Descriptive Statistics* in section 3.4.1, followed by the *Complex Network Analysis* (containing our chosen Complex Network metrics) in section 3.4.2. Then we discuss the selection of the most appropriate *Graph Visualisation Analysis* (including their respective visualisation and simulation algorithms) in section 3.4.3, and lastly the *Statistical Network Analysis* in the later section 3.5.7.

3.4.1 Descriptive Statistics

Since our data collection revealed a significant amount of *traceroute* hop observations, our first problem was to arrange the analysis towards a more manageable form. Hence, we described, summarised and indicated the data in a meaningful way. By describing the essential features of the upstream *traceroute* data using *Descriptive Statistics*, we were not only able to get a basic understanding of the collected *traceroute* observations and what the mapped reality revealed, but also to narrow down the number of observations on a per mobile broadband operator basis, which were of real interest for our single case study strategy.

The observations of interest were associated with the connections originating from the utilised Aircel, Bharti Airtel and Vodafone SIM cards as defined above. We obtained these per operator observations by utilising filtering mechanisms on our collected *traceroute* dataset (see section 3.3.4 below). Finally, we only described those insights that were feasible to analyse using the exploratory steps of analysis stated in section 3.5 below. This set the groundwork for further exploratory analysis by using *Complex Network*, *Graph Visualisation* and *Statistical Network Analysis*. Finally, only descriptions that were grounded in the collected raw data were reported.

3.4.2 Complex Network Analysis

Building on the findings of the *Descriptive Statistics* for the relevant observations from the collected *traceroute* dataset revealed the need to better understand the modelled reality of the upstream Internet market structure for the three Tamil Nadu mobile broadband operators.

Given the non-trivial nature of the upstream Internet market structure, we chose to study Complex Network metrics of the three Tamil Nadu mobile broadband operator networks. These metrics are frequently used to examine complex systems, which usually involve a large number of highly interconnected units of interest. Examples include works on Neural Networks, Biological Systems, Statistical Physics, the World Wide Web, or the Internet structure, amongst others (see e.g. Faloutsos, Faloutsos and Faloutsos, 1999; Strogatz, 2001; Boccaletti et al., 2006).

Therefore, *Complex Network Analysis* relates to a branch of Network Theory that aims to study non-trivial structural properties. The Internet periphery structure itself can thereby be explored at different levels of granularity, while the research landscape provides no

agreement on best practices of these granularities (see section 2.3.2). Huffaker, Fomenkov and Claffy (2016) provide one of the most comprehensive definitions. They argue that the Internet may be studied at (Glass) Fiber, IP address, Router, Points-of-Presence (artificial interface between connecting entities), Autonomous System and *Internet Service Provider* granularities. Here, based on the obtained *traceroute* data, the upstream Internet access market structure was analysed at two of these granularities, the Internet Protocol (IP) address and the Autonomous System (AS) granularity. This choice made particular sense considering that these granularities provide valuable insights into more economic and policy needs of the largely unregulated peering ecosystem amongst Autonomous Systems on the Internet rather than purely technical ones (Huffaker, Fomenkov and Claffy, 2016). Moreover, the *Internet Service Provider* granularity would have been of additional value but it is almost impossible to obtain data that relates Autonomous Systems (ASes) to *Internet Service Providers*, since any *Internet Service Provider* may manage or co-manage a multitude of Autonomous Systems. Furthermore, an Autonomous System may also represent a company that is associated with an *Internet Service Provider* as a single legal entity. Therefore, we consider the Autonomous System granularity as most valuable for effectively studying the upstream Internet market structure, while sometimes referring to their likely associated *Internet Service Providers*. Nevertheless, the Autonomous System granularity, or AS granularity, is only obtainable through first analysing the Internet Protocol (IP) granularity. The metrics to analyse the different granularities are Complex Network metrics, which usually relate to Sociology, Physics or Mathematics, more specifically to Graph Theory. The following section explains the chosen Complex Network metrics that were used to explore the upstream Internet market structure in the upcoming Chapters 4 and 5.

Choice of Complex Network Metrics

Here, we explain the relevant Complex Network metrics that were used for the *Complex Network Analysis*, the *Graph Visualisation Analysis* and the *Statistical Network Analysis* in this dissertation. Moreover, since these metrics are usually described through Physics or Sociology perspectives, it is vital to describe the specific applications of the chosen metrics for studying the upstream Internet structure. For this purpose, we expand on the usage of Network Analysis as introduced in the Literature Review (section 2.3). Therefore, this section aims to provide a shared understanding of applying the Complex Network metrics for the analysis of our collected *traceroute* data at both Internet Protocol and Autonomous System granularity.

First, we explain the general network metrics, followed by further relevant edge and vertex metrics used in the *Descriptive* and *Complex Network Analysis* in Chapter 4 (sections 4.1, 4.2 and 4.3). This section concludes by stating the Complex Network Models being employed through utilising graph visualisations, simulations and computations using *Graph Visualisation Analysis*.

General Metrics

Graphs Definition

For simplicity, let graph G of a network be defined as a collection of vertices V , representing either IP addresses or Autonomous System Numbers. These vertices connect to one another through edges E , whereas each edge represents a directed connectivity relationship between any pair of IP addresses or Autonomous System Numbers. Figure 3-4 below illustrates an example network graph G_C (see explanation of graph denotations on the following page), where vertex 1 is connected to vertex 2 through a directed edge and vertex 2 to vertex 3 through another directed edge. Hence, we see a one-sided directed connectivity (as applicable for *traceroute* observations).

Graph visualisation of a fictive network G_C



Key

- Vertex without label.
- Directed edge, linking a pair of vertices.

Figure 3-4: Example vertices and edges.

Each vertex represents an object in a network graph, whereas each edge accounts for a joint between a pair of two distinct vertices. More formally, graph G denotes as:

$$G = (V, E)$$

Our following network analysis also incorporates loops that represent vertices linking to themselves. We learned that this behaviour is especially common when analysing the networks at Autonomous System granularity where different IP addresses of the same IP address prefix connect to each other, representing internal routings between IP addresses in Autonomous Systems. All network graphs in this case study are considered as directed graphs, due to the directed nature of *traceroutes*. Again, this means that each edge represents a directed edge from one vertex to another one, except for where otherwise stated.

Degree

The *Degree*, $\deg(v_i)$ of a vertex v_i represents the total number of IP addresses or Autonomous Systems that are adjacent (joint by an edge, or relationship) to one IP address or Autonomous System represented by:

$$\deg(v_i)$$

where the maximum *Degree* of a network is denoted by $\Delta(G)$ and the minimum by $\delta(G)$. Vertices with a *Degree* of zero are isolated, while vertices with a *Degree* of '1' are leaf or end vertices. n represents the total number of vertices. Vertices with a *Degree* of $n - 1$ are dominating vertices. Therefore, in our Figure 3-4 above, vertex 1 would have a *Degree* of 1, vertex 2 a *Degree* of 2 and vertex 3 a *Degree* of 1. Consequently, $\Delta(G)$ is given by the *Degree* of vertex 2, while $\delta(G)$ is jointly given by the *Degree* of vertices 1 and 3.

Network Diameter

The *Network Diameter* represents the longest possible distance (longest shortest path), $\max_{v_i, v_j} \text{dist}(v_i, v_j)$ for any calculated shortest paths of two vertices, v_i and v_j in a general graph G (Harary, 1994, p.14), where $\text{dist}(v_i, v_j)$ represents the general graph G 's distance, and therefore represents the largest maximum number of hops from a *traceroute*'s source to its priori randomly-assigned destination. This represents the given data packet traversal, and hence IP address connectivity, through the Internet. While a disconnected network graph would have an infinite *Network Diameter* (Bliss and West, 2007), a shorter *Network Diameter* represents higher connectedness under lower longest possible distances. Therefore, longer *Network Diameters* are likely to have a negative impact on a mobile broadband operator's Quality of Service (QoS) that is delivered to their end-users in the Internet periphery. Similarly, a shorter *Network Diameter* may

indicate a positive impact on a mobile broadband operator's Quality of Service (QoS).

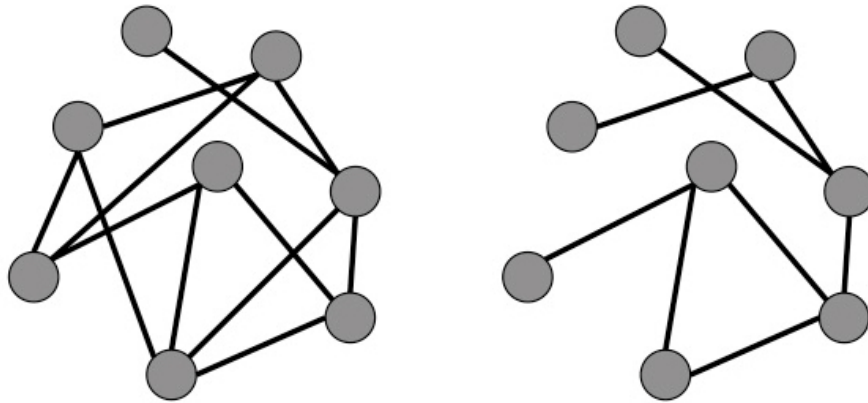
Network Density

The *Network Density* D in a general graph G is the ratio of the number of a networks' edges E over the total number of possible edges given by the binominal coefficient $\binom{N}{2}$, where *Network Density* D is calculated as:

$$D = \frac{2E}{N(N-1)}$$

The *Network Density* D for the network graphs of this case study describes the portion of existing connections over all the potential connections, meaning that a potential connection could exist between any two vertices v_i and v_j regardless of whether or not the connection actually exists. A network, where any vertices are directly connected, can be denoted as being perfectly dense. Hence, the example Network graph visualisation G_A in Figure 3-5 on the next page still employs a higher network density than the example network graph visualisation G_B , while not being perfectly dense connected.

Graph visualisation of fictive networks G_A (left) and G_B (right)



Key

- Vertex without label.
- Undirected edge, linking a pair of vertices.

Figure 3-5: Example network displaying Network Density.

Network Modularity

The *Network Modularity*, which measures strengths of a network division into clusters of associated IP address or Autonomous System vertices, captures the overall network structure. Figure 3-7 and Figure 3-8 on the following pages visualise such clusters in a given example network. Hence, networks with a high *Network Modularity* have more dense connections between vertices within clusters but rather sparse connections between vertices in different clusters. Analysing the *Network Modularity* helps to detect structures of community organisation, which might be an important measurement for the analysis of Autonomous System relationships under peering agreements or other business relationships.

After having discussed the general network metrics, next we introduce metrics that capture specific characteristics of the individual network's edges.

Edge-Metrics

Adjacency Matrices, In-Degree and Out-Degree

One may start to capture the elementary properties of the connectivity for each *traceroute* t by stating its *Adjacency Matrix* $A^t = [a_{ij}^t]$, where

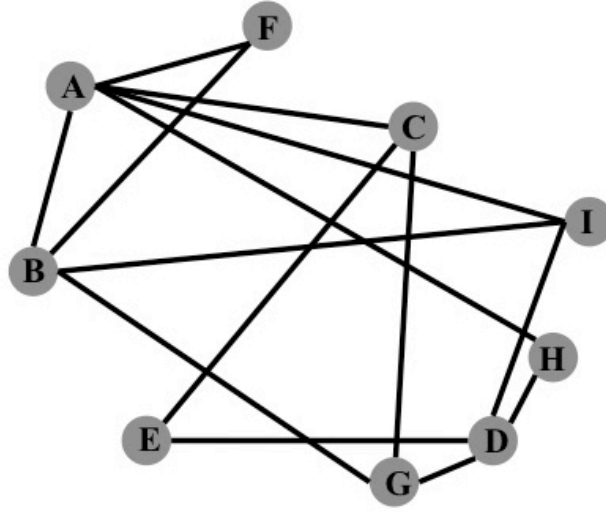
$$a_{ij}^t = \begin{cases} 1, & \text{if } \{i, j\} \in E, \\ 0, & \text{otherwise} \end{cases}$$

So that a_{ij}^t is non-zero for those entries whose row-columns indices correspond to vertices joined by a direct edge in the network G generated by the observed *traceroute* observations, and zeros for those that are not. Table 3-2 below represents a simple example *Adjacency Matrix* for a fictive network G_C with 9 vertices (A – I) and their respective undirected binary edges (1 for edge exists, zero for otherwise), linking the pairs of these vertices as stated in Table 3-2 below. Additionally, Figure 3-6 on the next page visualises this network graph of G_C .

Example Adjacency Matrix									
G	A	B	C	D	E	F	G	H	I
A	0	1	1	0	0	1	0	1	1
B	1	0	0	0	0	1	1	0	1
C	1	0	0	0	1	0	1	0	0
D	0	0	0	0	1	0	1	1	0
E	0	0	1	1	0	0	0	0	1
F	1	1	0	0	0	0	0	0	0
G	0	1	1	1	0	0	0	0	0
H	1	0	0	1	0	0	0	0	0
I	1	1	0	0	1	0	0	0	0

Table 3-2: Example Adjacency Matrix

Graph visualisation of a fictive network G_c



Key

● Vertex with label.

— Undirected edge, linking a pair of vertices.

Figure 3-6: Example network displaying an Adjacency Matrix.

Due to the directed nature of the collected *traceroute* observations, one may differentiate between the number of edges pointing towards a certain vertex, the vertex *In-Degree* d_v^{inde} , and the number of edges pointing away from a vertex towards the next or final one, the vertex *Out-Degree* d_v^{outd} . The denotation of the *Adjacency Matrix* therefore allows us to express the *In-Degree*, d_i^{inde} , and *Out-Degree*, d_i^{outd} , as being the directed connectivity of each vertex in a given *traceroute* t from the total number of observed *traceroutes* T originating from the three mobile broadband operators, where $t \in T$ (total number of *traceroutes*) as:

$$d_i^{outd} = \sum_j a_{ij}^t \quad \text{and} \quad d_j^{inde} = \sum_i a_{ij}^t$$

In Figure 3-4 above, vertex 1 would have an *In-Degree* of zero, $d_j^{inde} = 0$, but an *Out-Degree* of ‘1’, $d_i^{outd} = 1$. Vertex 2 would have a $d_j^{inde} = 1$ and a $d_i^{outd} = 1$. This is of particular relevance since the *traceroutes* of this case study represent directed connections between IP address vertices. Hence, our metrics profit greatly from taking the weights of such edges into consideration, revealing strongly connecting vertices. Given the example

Adjacency Matrix of the fictive network G_C , an *Adjacency Matrix* for our *traceroute* observations would indicate connections with varying weights between IP addresses. Therefore, an often-traversed connection or edge might, for example, obtain an *In-Degree*, $d_j^{inde} = 12$ and an *Out-Degree*, $d_i^{outde} = 12$. Again, our collected dataset covers 57,122 unique *traceroutes*, whereas each *traceroute* itself can be denoted by an *Adjacency Matrix*, A^t . The sum of all *traceroutes*' *Adjacency Matrices*, one per observed *traceroute*, t , for all $t \in T$ is a weighted network, or final matrix A , whereas $A = \sum_{t \in T} A^t$. The elements A_{ij} of the matrix A are non-negative numbers, showing how many times a given directed connection was observed between two IP addresses or Autonomous Systems in the set of all *traceroutes*, T , equivalent to the sum of binary observations A_{ij}^t for all possible *traceroutes* being $t \in T$. From the final matrix $= \sum_{t \in T} A^t$, one may derive the corresponding *Weighted In-* and *Out-Degrees* of the observed networks, a key element in the *Statistical Network Analysis* in Chapter 5 since it accounts for the above described weights of connectivity relationships, being:

$$d_i^{woutd} = \sum_{t \in T} \sum_j a_{ij}^t \quad \text{and} \quad d_j^{winde} = \sum_{t \in T} \sum_i a_{ij}^t$$

where the *Weighted Degree* of any vertex is given by the sum of the vertex's *Weighted In-* and *Weighted Out-Degrees*, as:

$$d_i^{wd} = d_i^{winde} + d_i^{woutd}$$

Path Lengths

The *Path Length* L_p of a *traceroute* path p helps to quantify the structural properties of any graph G by measuring typical separations between two vertices as global property. Assuming an undirected graph G , one can suggest the *Path Length* is a sequence of vertices such as:

$$L_p = (v_1, v_2, \dots, v_n) \in V$$

where v_p is adjacent to v_{p+1} for $1 \leq p < n$. Therefore, the *Path Length* L_p , ranging from v_1 to v_n has the length of n . Shorter *Path Lengths*, being used as connectivity metric, may be considered to facilitate a quicker transfer of information. Therefore, shorter *Path Lengths* might theoretically be valuable for reducing upstream connectivity costs while also improving an end-user's perceived Quality of Service (QoS).

Average Path Length

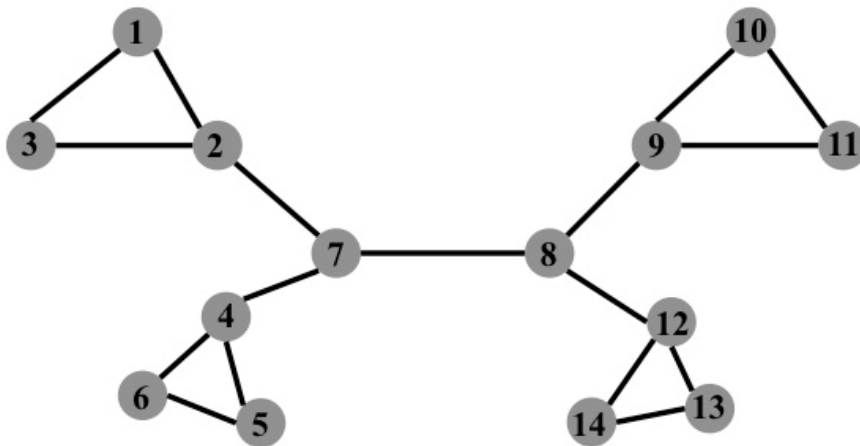
The *Average Path Length*, $Lavg_G$, can be obtained for an unweighted graph G by denoting the shortest distance between two vertices, v_i and v_j as $d(v_i, v_j)$, where v_i and $v_j \in V$. Assuming that the vertex v_j cannot be reached from vertex v_i , our *Average Path Length*, $Lavg_G$ can be denoted as:

$$Lavg_G = \frac{1}{V \cdot (V - 1)} \cdot \sum_{i \neq j} d(v_i, v_j)$$

Edge Betweenness

The *Edge Betweenness* is roughly defined by the number of shortest paths sp_{ij} going through an edge, e , linking a pair of vertices v_i and $v_j \in V$. The *Edge Betweenness* is related to the amount of traffic that any edge carries between a pair of Internet Protocol or Autonomous System vertices in a connection.

Graph visualisation of a fictive network G_D



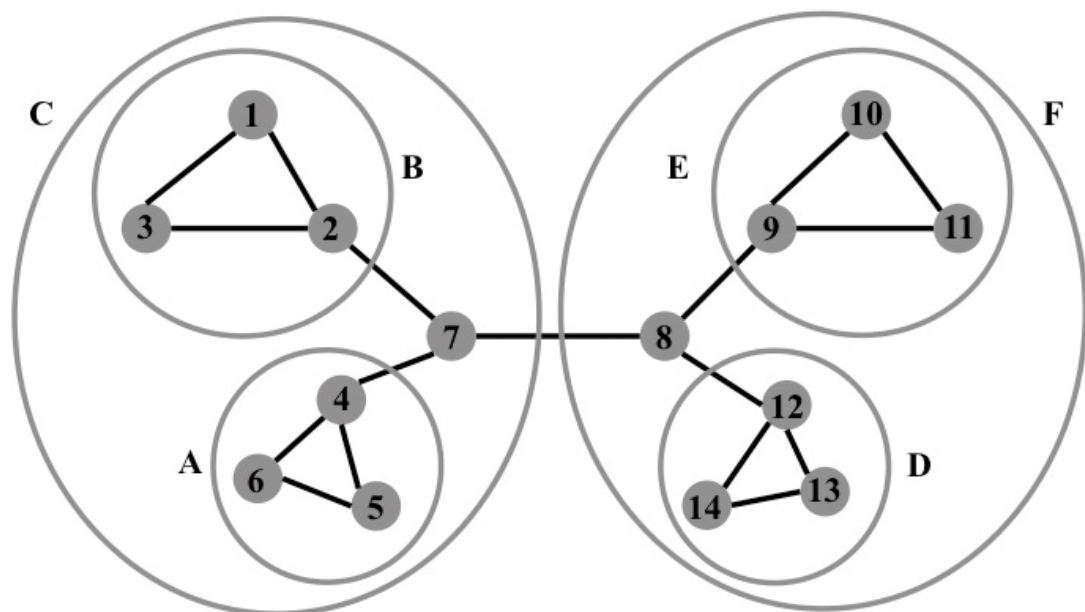
Key

- Vertex with label
- Undirected edge, linking a pair of vertices.

Figure 3-7: Example network displaying Edge Betweenness.

Considering the graph visualisation of an example network G_D in Figure 3-7, the *Edge Betweenness* for v_1 and v_2 would be '1', similar to a multitude of other vertices in this network. The *Edge Betweenness* for v_2 and v_7 would also be '1' but providing a total amount of flow it carries would be '33' (linked to v_2) x '11' (linked to v_7) = '33'. This provides structural implications. The graph visualisation of the fictive network G_D in Figure 3-7 is bi-partitioned, meaning that there are two almost separate sub-graphs 'C' and 'F' of the Network (visualised in Figure 3-8 below). Those vertices with a high *Edge Betweenness* carry the biggest load and occupy structural gatekeeping or congestion roles in a network. The vertices v_7 and v_8 linking the bi-partioned sub-graphs are therefore of strong importance for this example network. Here, vertex v_7 not only connects the vertex clusters 'A' and 'B' but also provides access to the vertex clusters 'D', 'E' and 'F'.

Graph visualisation of a fictive network G_D



Key

● Vertex with label.

— Undirected edge, linking a pair of vertices.

Figure 3-8: Example network displaying Neighbourhood Overlap.

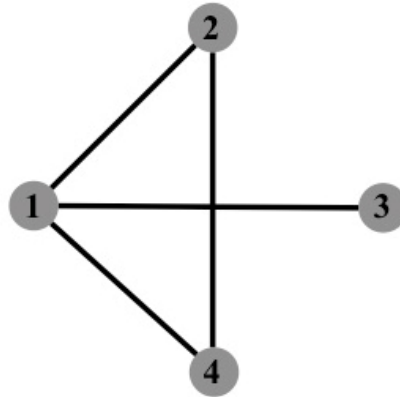
Neighbourhood Overlap

The *Neighbourhood Overlap* of an edge e connecting v_i and v_j covers the number of vertices that are neighbours of v_i and v_j over the number of vertices that are neighbours of at least one of both adjacent vertices, v_i or v_j . Hence, the *Neighbourhood Overlap* represents the intersection of the union of the neighbours. The key feature of the *Neighbourhood Overlap* is that the ratio is zero when the numerator is zero, meaning that the edge e is a local bridge (Easley and Kleinberg, 2010, p.47). Edges, and therefore any hops of a *traceroute* with a very small *Neighbourhood Overlap* have almost no other vertex in common. Therefore, the smaller the *Neighbourhood Overlap*, the more unique the connection. Similarly, the larger the *Neighbourhood Overlap*, the more frequent or common the connection. This is especially interesting when looking at the relationships between a set of Autonomous Systems vertices.

Embeddedness

The *Embeddedness* of an edge e in a given network is the number of common neighbours the two endpoints have (Easley and Kleinberg, 2010). By assuming a graph visualisation of the fictive network G_E with four vertices, v_1, v_2, v_3, v_4 in the following Figure 3-9, Vertex v_1 has an edge to the three vertices v_2, v_3 and v_4 while Vertex v_2 has edges to vertices v_1 and v_4 . The edge between vertex v_1 and vertex v_2 shows an *Embeddedness* of '1' since vertex v_1 and vertex v_2 have one common neighbour, v_4 .

Graph visualisation of a fictive network G_E



Key

● Vertex with label.

— Undirected edge, linking a pair of vertices.

Figure 3-9: Example network displaying *Embeddedness*.

Embeddedness is a social network theory metric frequently used in the analysis of sociological problems and can be seen as a property of structure (Kogut et al., 1992), in which vertices (such as social actors) behave and act as being embedded in concrete and ongoing systems of social relationships, relating to macro-level interests of markets and hierarchies (Granovetter, 1985, p.507). The *Embeddedness* is therefore an interesting measurement for the analysis of potential cooperation among Internet Protocol addresses or Autonomous Systems since those are most likely manually added to the routing tables for connectivity purposes.

Vertex-Metrics

Clustering Coefficient

The *Clustering Coefficient*, \mathbb{C}_i , or network transitivity, of each vertex, v_i in a directed graph is the ratio between the existing edges, E , amongst all other vertices, V , being connected to this same vertex v_i (in our application either an IP address or an Autonomous System vertex, depending on the level of granularity adopted) over the maximum number of potential interconnections (Boccaletti et al., 2006, p.10). Therefore, the measure of *Clustering Coefficient* captures mutual interconnections of direct neighbour-vertices k_i , of any vertex v_i , whereas $v_i \in V$ measures the probability that any peers (neighbours) of a vertex are connected between themselves, referring to a *Small-*

World Network phenomenon (Watts & Strogatz, 1998). This metric is of particular importance for Autonomous System vertices and the internetworking of the set of Autonomous System vertices, including their neighbours and therefore helps to measure interconnection and network structuring (Vázquez, Pastor-Satorras and Vespignani, 2002, p.3). We set the *Clustering Coefficient*, \mathbb{C}_i as:

$$\mathbb{C}_i = \frac{2|\{e_{lk}: v_l, v_k \in N_i, e_{lk} \in E\}|}{k_i(k_i - 1)}$$

where N_i is the set of vertexes directly connected to vertex v_i (or v_i 's neighbours). The *Average Clustering Coefficient* is the mean value of the individual *Clustering Coefficients* \mathbb{C}_i .

Weighted Clustering Coefficient

When studying Complex Networks, Barrat et al. (2004) have shown that the *Weighted Clustering Coefficient*, \mathbb{C}_i^w , is needed to explore structural organisation and structural network information in non-trivial systems such as the Internet. Therefore, the *Weighted Clustering Coefficient* gives a measure of local cohesiveness, which takes the amount of vertex interaction intensity, and this Internet traffic for each local triplet, into consideration. The *Weighted Clustering Coefficient* can be denoted as:

$$\mathbb{C}_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{(w_{i,j} + w_{i,h})}{2} a_{i,j} a_{i,h} a_{j,h}$$

where $a_{i,j}$ is the element of an *Adjacency Matrix*; in row i and column j , k_i is the degree of vertex i ; $s_i(k_i - 1)$ represents the normalisation factor that accounts for the weight of each edges multiplied to the maximum number of possible edge triangles that participate. This ensures that $0 \leq \mathbb{C}_i^w \leq 1$. Moreover, \mathbb{C}_i^w recovers the topological *Clustering Coefficient* as long as $w_{i,j} = \text{const.}$

Centrality-Metrics

Degree Centrality

This metric simply refers to the number of edges that a vertex, v_i , has. Moreover, it assumes linearity, meaning that if vertex v_i has twice as many connected vertices than vertex v_j , then vertex v_i is twice as important.

$$c_i^d = \sum_{j:j \neq i} v_{i,j}$$

In a directed graph, *Degree Centrality* is usually measured through *In-Degree Centrality* for incoming connections to a vertex:

$$c_i^{inde} = \sum_{j:j \neq i} v_{j,i}$$

and *Out-Degree Centrality* for a vertex' outgoing connection.

$$c_i^{outde} = \sum_{j:j \neq i} v_{i,j}$$

In relation to an Autonomous System network, vertices with a lower *Degree Centrality* would be more peripheral in the network, whereas Autonomous Systems with a higher *Degree Centrality* would be more central.

Betweenness Centrality

The *Betweenness Centrality* is used to capture a degree of unavoidability of a given vertex, showing the proportion of times that a vertex appears on the shortest paths between any other two vertices, or how many pairs of vertices would have to go through a certain vertex in order to reach one another in a minimum number of hops (Freeman, 1977). When stating statistical theory of Internet exploration, Dall'Asta et al. (2005) note that the *Betweenness Centrality* of a vertex v_i , $C_B(v_i)$, covers many topological properties very well. According to Kolaczyk (2009), *Betweenness Centrality* summarises the extent to which a vertex is located between other pairs of vertices. One can therefore denote *Betweenness Centrality* by:

$$C_B(v_i) = \sum_{i \neq j} sp_{ij}(v_i) / sp_{ij}$$

where $sp_{ij}(v)$ represents the number of shortest paths connecting i and j , passing through v and sp_{ij} , the total number of shortest paths. Here, the *Betweenness Centrality* assumes that communication between Autonomous System vertices always follow the shortest paths. This aspect is unlikely to be applicable in real-world networks just as those mobile broadband ones in this dissertation. Based on this logic, we therefore take a different perspective given a statement of Dall'Asta et al. (2005). Nevertheless, the metric will still be analysed due to its interesting vertex location properties.

Closeness Centrality

Closeness Centrality quantifies the importance of a vertex based on the inverse of the average distance between a vertex and all the other vertices of a network (Freeman, 1978; Wassermann and Faust, 1995). The classic *Closeness Centrality* is proposed by Bavelas (1950), Beauchamp (1965) and Sabidussi (1966) as:

$$c_c(j) = \sum_i \frac{1}{sp_{ij}(v)}$$

where $sp_{ij}(v)$ represents the shortest path connecting the vertices i and j in a given Network. *Closeness Centrality* shows how close a vertex is to the other network vertices (Kolaczyk, 2009). Just like above, the *Closeness Centrality* assumes the existence of shortest paths between Autonomous Systems, which does not represent real-world network functioning of our mobile broadband operator networks.

Eigenvector Centrality

The *Eigenvector Centrality* is a measurement of vertex influence in a given network. Therefore, the metric assigns scores for given vertices. The score is higher for those vertices that connect to high-scoring ones rather than low-scoring vertices in the overall network. Hence, high-scoring vertices contribute more to a scoring than low-scoring ones. Let us assume a given graph $G = (V, E)$ with $|V|$ vertices and $A^t = [a_{ij}^t]$ being the graph G 's *Adjacency Matrix* again where

$$a_{ij}^t = \begin{cases} 1, & \text{if } \{i, j\} \in E, \\ 0, & \text{otherwise} \end{cases}$$

then the relative Eigenvector Centrality score c_i^e can be defined as

$$c_i^e = \frac{1}{\lambda} \sum_{j \in M(i)} c_j^e$$

where $M(i)$ represents the neighbors of the vertex i and where the eigenvalues λ are constant. This can be denoted as the eigenvector equation in vector notation as:

$$Ac^e = \lambda c^e$$

However, this might reveal a multitude of eigenvalues λ for which a non-zero eigenvector solution exists. Hence, we set the requirement that all entries in the eigenvector are non-zero (Perron-Frobenius Theorem), meaning that only the greatest eigenvalue results in

a desired measure. Next, the i^{th} component of the related eigenvector provides the scoring to vertex i . The eigenvector can only be defined up to a certain factor. Hence, vertices can only obtain a ratio of the centralities. Defining the absolute score comes after normalising the eigenvector that the sum of all vertices is '1'. Here, an Autonomous System vertex i with a larger *Eigenvector Centrality* c_i^e would therefore show the strongest influence in a given network.

Complex Network Models

Complex Network models refer to non-trivial structural network properties that occur in the modelling and simulation of real-world network graph structures. In the following paragraphs, we aim to explain the general differences between a number of such Complex Network models using the specific algorithms employed.

Barabási-Albert Model

The Barabási-Albert Model is a random network model to simulate 'rich-get-richer' effects, called preferential attachments. Here, we explain the Barabási-Albert dynamic network procedures that simulate alternative scenarios of network growth emergence. A vertex v_i is therefore more likely to attach to vertices that have higher *Degrees*. Vertices with a *Degree* of '0' remain disconnected from the rest of the network whereas the initial network begins with a number of vertices, m_o . If $m_o \geq 2$ then the *Degree* of each vertex should be '1'. The BA Model therefore constantly adds new vertices under the rule mentioned above. The probability p_{v_i} that vertices are connected to a vertex i is denoted as:

$$p_{v_i} = \frac{\deg(v_i)}{\sum_{v_j} \deg(v_j)}$$

where $\deg(v_i)$ represents the *Degree* of the vertex v_i (shown above) and the sum is made over all pre-existing vertices v_j . Therefore, strongly-linked vertices, here connectivity-important Autonomous Systems, accumulate quickly into hubs since they have a stronger preference attached to them. The underlying *Degree* distribution is *Scale-Free* and can be denoted as a *power-law degree distribution*:

$$P(\deg) \sim \deg^{-3}$$

Section 2.3.2 in the Literature Review revealed, that the Internet shows, at different levels of granularities, these *power-law degree distributions*. Hence, we consider the Barabási-

Albert Model for network growth as very valuable to simulate network growth emergence in mobile broadband operator networks (Barabási and Albert, 2002). This might be valuable to study connectivity simulations.

3.4.3 Graph Visualisation Analysis

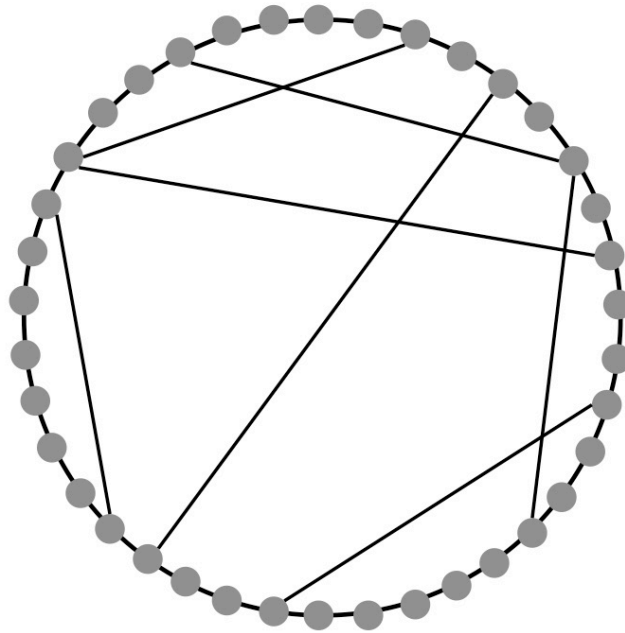
The *Graph Visualisation Analysis* uses graph-drawing techniques to visualise network diagrams consisting of their vertices and edges linking these vertices in two-dimensional Euclidean spaces. *Graph Visualisation Analysis* is closely related to Graph and Network Theory. Moreover, there is a great research interest in visualising the structure of the Internet. The La Jolla, CA – based Center for Applied Internet Data Analysis has engaged in Internet visualisation efforts since the year 2000 (CAIDA, 2015). A great number of possible network graph visualisations are covered as the ‘Internet Topology Zoo’ at the University of Adelaide (2016). A common framework for visualising and analysing the structure of the Internet graphs are the different granularities that can be adopted, as stated in section 2.3.4 above. Given the visualisation efforts, it is important to have a clear understanding of the graph visualisations that are worth analysing to gain a structural understanding. We consider the *Graph Visualisation Analysis* as an important step in further exploring our findings from the *Descriptive* and *Complex Network Analysis* in Chapter 4. Visualising the mobile broadband operator network graphs might yield additional insights for understanding key structural properties. The following section below discusses the reasoning for choosing certain graph visualisation and simulation algorithms and our choice of distinctive visualisation layouts over others.

Graph Visualisation Algorithms

Kleinberg Small-World Network Model

A *Small-World Network* model refers to a mathematical representation where most vertices in a given network graph are not neighbours of one another but instead, the neighbours of given network vertices are likely to be neighbours of each other so that most network vertices are reachable from every other vertex given a small number of hops (or steps). This means that relatively short paths exist between any two vertices in a given network (Watts and Strogatz, 1998, p.440). Moreover, the typical distance between randomly chosen vertices in a *Small-World Network* grows proportionally to the logarithm of the number of all vertices in a given network, while *Clustering Coefficients* in *Small-World Networks* are naturally, given the above explained effect, large (Watts and Strogatz, 1998, p.442).

A suitable algorithm to study the *Small-World Network* effect is Kleinberg's *Small-World Network* model (Kleinberg, 2000). This model uses so-called greedy routing algorithms. This means, in our context, that an IP address vertex in a given *traceroute* path could choose the next vertex it believes to be closest to the chosen destination, based on the *Small-World Network* effects (Kleinberg, 2000). This effect in the Kleinberg Model is achieved by adding long-range edges to the network, which tend to favour vertices that are closer in distance (not geo-distance), rather than farther. The *Small-World Network* phenomena are well visualised in using graph visualisation algorithms such as the *Layered Layout* by Kuchar (2012), which visualises vertices in different layers, depending on the values of a chosen vertex property.

Small-World Network graph visualisation**Key**

- Vertex without label.
- Undirected edge, linking a pair of vertices.

Figure 3-10: Small-World Network graph visualisation.

Barabási-Albert (BA) Scale-Free Model

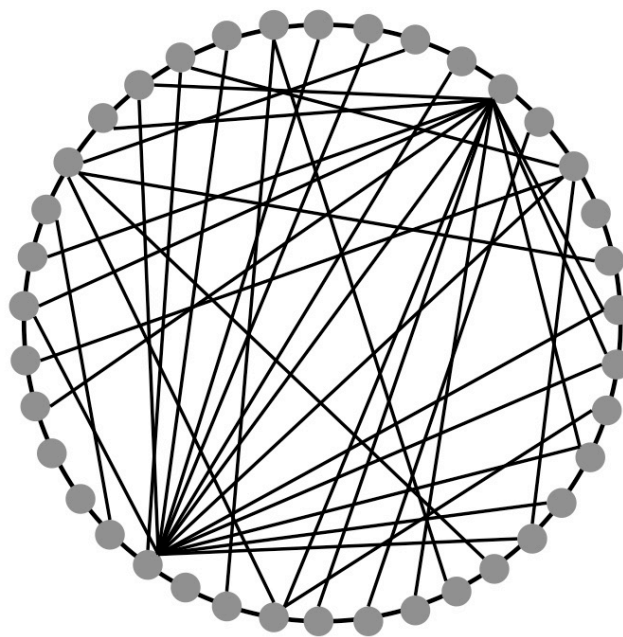
Scale-Free Network models are those whose *Degree* distribution follows *Pareto* or *power-law degree distributions*. Albert, Jeong and Barabási (1999) find that the World Wide Web follows such a *power-law degree distribution* and hence *Scale-Free Network* properties. The Barabási-Albert dynamic network algorithm simulates alternative scenarios of network growth emergence (Barabási and Albert, 2002). These growth features are achieved through preferential attachment, referred to as ‘rich-get-richer’ effects given the *power-law degree distribution*. Barabási Labs (2013) refer to three distinct Barabási-Albert *Scale-Free Network* models:

- Standard Model with vertex growth and preferential attachment to edges.
- Model A with vertex growth and uniform attachment of edges.
- Model B without vertex growth but preferential attachment to edges.

While Onnela et al. (2007) use the Barabási-Albert *Scale-Free Network* models to uncover the structure and tie strength in mobile communication networks, Faloutsos,

Faloutsos and Faloutsos (1999) believe that the Internet has a *power-law degree distribution*, which is criticised by a number of researchers including Willinger, Alderson and Doyle (2009) and Willinger and Roughan (2013). Nevertheless, we believe that the Barabási-Albert *Scale-Free Network* models are a good simulator choice to explore scenarios of network growth emergence for *traceroute*-based connectivity. Hence, each of the three Barabási-Albert *Scale-Free Network* models were utilised to study the three mobile broadband operator network graph visualisations.

Scale-Free Network graph visualisation



Key

- Vertex without label.
- Undirected edge, linking a pair of vertices.

Figure 3-11: Scale-Free Network graph visualisation.

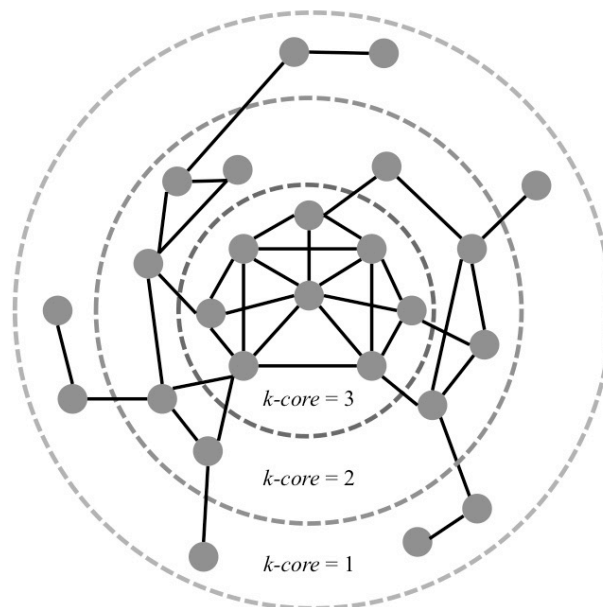
Scale-Free Network models are often visualised in using so-called force-directed Layouts. We chose two of these Layouts given their specific properties (a more comprehensive description of these Layouts is provided in the Literature Review, see section 2.3.4). Here, the *Force Atlas 2 Layout* is considered suitable for visualising *Scale-Free Networks* with between 10 and 10,000 vertices, which well suited to our *traceroute* observations (Jacomy et al., 2014). The *Force Atlas 2 Layout* incorporates a force-directed algorithm, which allows to place vertices in a two-dimensional space without crossing edges too much between the pairs of vertices, capturing structural properties of

a given network. Second, the *Fruchterman-Reingold* (1991) graph visualisation Layout is, just like the *Force Atlas 2 Layout*, a force-directed layout algorithm.

k-core decomposition

The *k-core decomposition* algorithm helps to study hierarchical properties of large scale networks through identifying particular subsets of a given network (Alvarez-Hamelin et al., 2005a, p.22), while being usually employed in biological settings to analyse and predict protein interactions (Seidmann, 1983; Alvarez-Hamelin et al., 2005b). The algorithm divides networks into different subsets, called *k-cores*. Therefore, the *k-core decomposition* focuses on the network regions with increasing centrality and connectedness. More central *k-cores* are, therefore, inhabiting more densely interconnected network vertices as Figure 3-12 below illustrates (through *k-cores* 1 – 3).

k-core decomposition graph visualisation



Key

- Vertex without label.
- Undirected edge, linking a pair of vertices.

Figure 3-12: *k-core decomposition graph visualisation*.

According to Alvarez-Hamelin et al. (2005b; 2008), the *k-core decomposition* allows the finding of connectivity paths with specific Quality of Service (QoS), especially when studying models of the Internet at Autonomous System granularity. Hence, the *k-core decomposition* seemed to be a very applicable modelling approach to analyse the key IP

address and Autonomous System vertices and hence connectedness regions of potential bottlenecks in the upstream Internet market structure, originating from the three Tamil Nadu mobile broadband operators of interest.

3.5 Exploratory Steps of Analysis

This section covers all exploratory iterations that we used in following the Dewey (1941) ‘warranted assertions’ inquiry process as explained in the sections 3.2 above (see also Boyles, 2006). Our choice of inquiry iteration steps incorporated our reflections on previous actions and beliefs while always keeping in mind the research aims and objectives and our abducted *Working Hypotheses* (see section 2.5) as end goals. For each subsequent inquiry iteration, we reconsidered the nature of the problem whereas our analytical steps were pragmatically suggested solutions, followed by taking and reporting analytical actions. The following inquiry iterations were employed throughout Chapter 4 and Chapter 5. In detail, Chapter 4 covered the Iterations 1 to 6, whereas Chapter 5 thoroughly covered Iteration 7. We start explaining these iteration steps by mentioning the underlying reasoning at the beginning of each section.

- Iteration 1: Descriptive Network Analysis (section 4.1).
- Iteration 2: Complex Network Analysis (IP granularity), (section 4.2).
- Iteration 3: Graph Visualisation Analysis (IP granularity), (section 4.2.6).
- Iteration 4: Complex Network Analysis (AS granularity), (section 4.3).
- Iteration 5: Graph Visualisation Analysis (AS granularity), (section 4.3.7).
- Iteration 6: Autonomous System Relationships (AS granularity), (section 4.4).
- Iteration 7: Statistical Network Analysis (AS granularity), (Chapter 5).

3.5.1 Iteration 1: Descriptive Analysis

The aim of the *Descriptive Analysis* of the Iteration 1 was to first gain a feeling and understanding of the collected 731,200 *traceroute* observations from our active Internet periphery measurements, originating from the three Tamil Nadu mobile broadband operators of interest. This step was necessary in order to determine the upcoming iterations. The outcomes of Iteration 1 were reported in section 4.1.

Hence, as a first step for quantitatively exploring the features of the collected *traceroute* observations, we had to get the *.txt *traceroute*’s raw data as well as the readme files from the Portolan (2015) Network Tools administrators (see section 3.3.4 above). Next,

we opened the *traceroute* raw data file into the spreadsheet analysis tool Excel. For further usage, we added a header row, where the columns were named according to the information provided by the Portolan (2015) Network Tools Administrator and saved the respective file as ‘india_traceroute_2015_03.xlsx’. The header descriptions provided a clearer view of the obtained *traceroute* raw dataset. Next, we counted the total number of observations per data collection days by filtering the days containing the ‘YYYY-MM-DD’ days of interest in the timestamp column (see Figure 4-1). To get a better sense of the amount of *traceroutes*, we counted the number of observed *traceroutes* by copying the *traceroute* identifier column and reported them into a new sheet named ‘Number of Traceroutes’ and removed existing duplicates. This allowed us to calculate the average number of IP observations per *traceroute*. Next, we estimated the number of IP source addresses and IP hop addresses by copying the column of IP source addresses and pasting them into a new sheet called ‘Number IP source’ and again, removed existing duplicates. We followed the same approach for the number of IP hop addresses and denoted this sheet ‘Number IP hop’. Then we copied both the number of unique IP source addresses and the unique IP hop addresses into a new column in a new sheet called ‘unique IP addresses’. This allowed us to get a sense for all IP addresses involved in the 731,200 *traceroute* hop observations. By filtering the campaign identifier by Autonomous System Numbers, we obtained the number of *traceroute* observations that originated from each Autonomous System as well as the percentage of observations compared to the total observations. This was a crucial step since only those hop observations that were originating from this case studies’ three Tamil Nadu mobile broadband operators were of interest. To find out which of the campaign identifiers were of relevance, we utilised the Autonomous System Number lookup feature in the Hurricane Electric (2016) BGP-Toolkit and double-checked the results with UltraTools (2016) and Team Cymru (2016). This allowed us to obtain the AS Numbers and for the relevant campaign identifiers of the three mobile broadband operators of interest. To measure if these observations were enough narrowed-down, we further filtered the operating system of the device, only focusing on those observations that originated from the utilised Android smartphones (see section 3.3.4). The obtained *traceroute* hop observations represented those per operator observations of interest. Based on these observations, we calculated the average number of unique *traceroute* hop and source observations per total *traceroute* observation originating from a specific Autonomous System. Furthermore, by obtaining the range for the lowest and highest number of the Round-Trip-Time (RTT) column, we were able to

shape an initial view of the Quality of Service (QoS) that an end-user might experience. This range was subsequently applied to all observations originating from our specific mobile broadband operators and added by a calculation of the variance between the lowest and highest Round-Trip-Times. Finally, we analysed and reported those *traceroute* observations originating from a particular Autonomous System Number that contained skips in the *traceroute*, an indicator for packet loss, again a potential indicator an end-user's perceived Quality of Service (QoS).

3.5.2 Iteration 2: Complex Network Analysis (IP)

Given the general understanding of the obtained *traceroute* observations from Iteration 1 above, Iteration 2 aimed to gain an initial understanding of upstream Internet market structure of the three Tamil Nadu mobile broadband operator networks using *Complex Network Analysis* of the data at Internet Protocol (IP) granularity. The outcomes of Iteration 2 were reported in Chapter 4.

To explore the collected *traceroute* hop observations at IP granularity, we utilised the Open Source Network Exploration Tool Gephi (2016). First, we opened our 'india_traceroute_2015_03.xlsx' raw data file in Excel, containing the collected *traceroute* hop observations from our active Internet periphery measurements using Portolan (2015). Unlike before, we deleted all columns except the source IP and hop IP address ones. Next, we added a header row and named the source IP address column 'Source' and the hop IP address column 'Target', a prerequisite for importing edge-tables into Gephi (2016). The file was then saved twice, once as 'complete_IP_import_for_gephi.xlsx' and then as 'complete_IP_import_for_gephi.csv' for the following import into Gephi (2016).

Once saved, we started Gephi (2016), created a new project and imported the elaborated *.csv file. Since our *traceroute* raw data comprised of connections between IP addresses, we imported the file as an edges-table, rather than a vertex-table, resulting in a directed graph where IP addresses were linked to their neighbouring IP addresses. Once the file was imported, we were able to calculate the relevant Complex Network metrics by using Gephi (2016)'s statistics functions. Next, we saved the file with the calculated metrics as 'complete_IP.gephi'. To calculate the *Weighted Clustering Coefficient*, we first had to download and install the Complex Generators plugin, generated by Bartosiak (2012), which was available on the Gephi Marketplace. Once the relevant Complex Network

metrics were calculated, we exported the resulting data table (including the calculated metrics) as ‘complete_IP_after_gephi.csv’. Next, we separated the *traceroute* hop observations of interest (starting from the three relevant mobile broadband operators) from the total number of observations. For this purpose, we again opened the raw data file in Excel and saved it as ‘Raw_Data_IP_Separation.xlsx’. Instead of deleting the same columns as before, we kept the source IP address, the hop IP address and additionally the campaign identifier columns. This allowed us to separate those *traceroute* hop observations that originated from the three mobile broadband providers of interest. Therefore, we filtered the campaign identifier column by ‘WORLDin55831’ for Aircel, ‘WORLDin45609’ for Bharti Airtel and ‘WORLDin38266’ for Vodafone. The numbering represented their respective Autonomous System Numbers, identified through Hurricane Electric (2016) and filtered by the data collection times and locations, as further described in Chapter 4. Each of the filtered source and hop IP addresses were then saved in separate sheets, named with the respective Autonomous System Number and operator name. Next, we added a header row again while naming the source IP address column ‘Source’ and the hop IP address column ‘Target’. The obtained files were then exported as a *.csv file, while the AS Number in the naming referred to one of the three mobile broadband providers of interest, namely ‘AS55831’ for Aircel, ‘AS45609’ for Bharti Airtel and ‘AS38266’ for Vodafone. Our analysis file was saved as ‘Raw_Data_IP_Separation.xlsx’. One at a time, each of the three generated *.csv files were then imported as an edge-table into Gephi (2016). Once imported, we followed the previous steps to calculate the relevant Complex Network metrics (see above) and reported our findings in Chapter 4. To obtain the vertex strength distributions of the three operator networks (utilised to capture the *power-law degree distributions*), we first exported the *Degree* column from Gephi (2016) and named the file as ‘operatorname_degree_distribution.csv’. To generate the respective *Degree* distribution plots, we used the Statistical Computing Tool R (2016). Here, the obtained *.csv files were transformed to *.txt ones by simply renaming them. Once saved, we opened R (2016) and computed the plot as the following script example for Aircel, see Appendices.

The generated *power-law degree distribution* plots (see script in Appendices) were then reported at the appropriate place in Chapter 4 (section 4.2.4). To further explore the obtained results, the following inquiry Iteration 3 looked at the *Graph Visualisation Analysis* of our three Tamil Nadu mobile broadband operator networks.

3.5.3 Iteration 3: Graph Visualisation Analysis (IP)

The above Iteration 2 provided us with initial structural insights for the three mobile broadband operator networks at IP granularity, while also indicating, based on the general, edge and vertex metrics, that the structuring likely followed a *Scale-Free Network* model. As a result of this learning, we wanted to know more about the structural properties of the three operator networks, while also gaining insight into connectivity importance of certain vertices that might indicate structural bottlenecks (given the apparent *Scale-Free Network* nature) for providing upstream internetworking features of the three mobile broadband operator networks. Therefore, exploring the first two network models (*Small-World* and *Scale-Free*) of the *Graph Visualisation Analysis* (and simulation of *Scale-Free Network* models) at IP granularity in this Iteration 3 aimed to better understand the structural properties of the three mobile broadband operator networks. The *Small-World Network* features were therefore analysed using the Kleinberg (2000) algorithm (see section 3.4.3 above), while the Barabási-Albert Models (Standard Model, Model A and Model B, see Barabási et al. (2013)) were used to simulate the *Scale-Free Network* nature of the operator networks. Computing the *k-core decomposition* algorithm and graph visualisation using R (2016) upon the work of Alvarez-Hamelin et al. (2006) was then chosen to reveal those vertices, indicating potential structural internetworking bottlenecks. Hence, this section starts by describing the steps to generate the *Small-World Network* Model below, followed by the Barabási-Albert *Scale-Free Network* Models and lastly the *k-core decomposition*. The outcomes of Iteration 3 were reported in section 4.2.6.

Small-World Network Model

To generate the Small-World graph visualisations, we first re-opened Gephi (2016), activated the Complex Generators plugin and created a new project. Next, we imported our previously generated raw data *.csv files as edge tables (e.g. ‘AS38266_for_gephi.csv’ for Vodafone) and started the Kleinberg algorithm by following ‘File > Generate > Kleinberg Small World Model’ in Gephi (2016). The graph visualisation layout was set at the *Layered Layout* by Kuchan (2012), which we had to download (from the Gephi Marketplace) and install prior to usage. Before visualising the graph, we computed the *Weighted Average Clustering Coefficient* measurements in the statistics section of Gephi (2016). Next, we chose the *Weighted Average Clustering Coefficient* as distance parameter in the *Layered Layout* graph visualisation. Once the graph visualisation was generated, we coloured the visualisation background ‘white’ and

the edges ‘blue’ and exported the resulting visualisations as *.png files while saving the models (e.g. as ‘AS38266_Kleinberg_blue.gephi’ for Vodafone). Lastly, we reported our findings alongside the utilised graph visualisation parameters. These were set in the Open Source Network Exploration Tool Gephi (2016) as:

- The calculated Weighted Clustering Coefficient.
- Layer Distance of 1250.
- Edge-thickness of 0.5, whereas due to readability purposes, we visualised the G_{Vodafone} with a smaller edge-thickness of 0.25.
- Size of lattice: 10.
- Lattice distance to local contacts: 2.
- Long range contacts: 2
- Clustering exponent: 0.
- ‘Black’ vertex colouring.
- ‘Light blue’ edge colouring.

Scale-Free Network Model

Obtaining the Barabási-Albert Scale-Free graph visualisations followed a somewhat similar approach. Here, we opened Gephi (2016) again and activated the Complex Generators plugin. Next, we started and saved a new project as *.gephi file (e.g. ‘AS55831_BAModel.gephi’ for Aircel) and imported the raw data *.csv file (e.g. ‘AS38266_for_gephi.csv’ for Vodafone) as edge table. We started the BA Standard Model algorithm by following ‘File > Generate > Barabási Albert Scale-Free Model’. Prior to this generation, we had to obtain the number of unique vertices for the graph. This information is found by opening ‘Window > Context’. The number of unique vertices was then included as ‘N Number of nodes in generated network’ (nodes is a synonym for vertices) in the settings of the Barabási-Albert *Scale-Free Network* algorithm. ‘M, the number of edges coming with every new node’ and ‘m0, number of nodes at the start time’ remained at ‘1’. Furthermore, we ticked the box to consider the existing vertices, representing the existing IP addresses or Autonomous Systems in the given three mobile broadband operator networks. Once the algorithm event finished the calculations, we generated the graph visualisations using the *Force Atlas 2 Layout*. To make the visualisation more readable, we made use of a specific set of graph layout parameters. Moreover, we changed the visualisation background colour again to ‘white’ and the colour of the edges to ‘blue’. Once the graph visualisations were generated, we

explored its components by setting a ‘k-core parameter’ as ‘topology query’. The respective graph visualisations were then exported as *.png files (e.g. ‘AS55831_BAModel_blue.png’ for Airtel). Next, we followed similar steps to generate the graph visualisations of the Barabási-Albert Model A and the Barabási-Albert Model B. Hence, we created and saved two new projects as *.gephi files (e.g. ‘AS45609_BAModel_nogrowth_blue.gephi’ and ‘AS45609_BAModel_uniformattachment_blue.gephi’ for Bharti Airtel).

We then launched the BA-model algorithm without growth by following ‘File > Generate > Barabási Albert Scale-Free Model B (no growth)’ and the BA-model algorithm without preferential attachment by following ‘File > Generate > Barabási Albert Scale-Free Model A (uniform attachment)’. For both models, we chose the number of unique vertices for ‘N Number of nodes in generated network’, obtained as stated above. Once the networks were generated, we again utilised the *Force Atlas 2 Layout* to visualise the generated graphs and exported the files in the *.png formats (e.g. ‘AS45609_BAModel_nogrowth_blue.png’ and ‘AS45609_BAModel_uniformattachment_blue.png’ for Bharti Airtel). Alongside their respective descriptions, the graph visualisations were then reported in section 4.2.6 of Chapter 4, where we also stated the following visualisation parameters to ensure comparability between the mobile broadband operator graph visualisations:

- Edge weight of 1 represents a normal edge influence.
- Fixed visualisation scale of 20 provides the graph visualisation with less repulsion.
- Normal gravity attraction of 1 assures that vertices are not leaving the two-dimensional Euclidean space.
- ‘Black’ vertex colouring.
- ‘Light-blue’ edge colouring.

k-core decomposition

Due to a lack of a suitable Gephi (2016) plugins, we modelled the computation and visualisation of the *k-core decomposition* as proposed by Alvarez-Hamelin et al. (2008) by using the Statistical Computing Tool R (2016). Hence, we installed the Network Analysis and Visualisation package ‘igraph’ by entering ‘>install.packages(“igraph”)’ in the R (2016) console. Once this package was installed, we wrote a R-script (see script in

Appendices), on the basis of Casas-Roma (2015), for computing and visualising the three mobile broadband operator networks' *k-core decompositions*. For readability purposes, we included comments (marked with a hashtag) in the script (see script in Appendices) below. However, this script only functions once the imported *.csv is cleaned from any header rows and identifier columns. These files were saved e.g. as 'AS45609_Bharti_for_R.csv' for Bharti Airtel. Having imported the respective *.csv files, R (2016) then calculated the *k-core decomposition* utilising the elaborated script, a task which needed significant computing resources. The resulting graph visualisations for the three Tamil Nadu mobile broadband operator networks were then exported and saved (e.g. as 'BhartiAirtel_kcore_decomposition.png' for Bharti Airtel). The exploration of the findings, alongside a comparison with the previous graph visualisations, was then also reported in section 4.2.6 of Chapter 4.

3.5.4 Iteration 4: Complex Network Analysis (AS)

The previous Iterations 1, 2 and 3 at Internet Protocol granularity above indicated that a *Complex Network Analysis* at Autonomous System (AS) granularity would bring additional value to reveal the structural importance of certain internetworking Autonomous Systems (rather than their lower granularity Internet Protocol addresses) in the upstream Internet market structure. Hence, after transforming our collected *traceroute* observations from Internet Protocol to Autonomous System granularity, this Iteration 4 aimed to re-calculate the general, edge and vertex metrics as calculated in Iteration 2.

Preparing the collected *traceroute* hop observations dataset for the analysis at Autonomous System required a transformation of the Internet Protocol (IP) addresses observations to their associated Autonomous System Numbers. Hence, we first downloaded the Maxmind (2015) Geo IP2 database for this purpose. This database is one of the most comprehensive collections of IP address ranges that any Autonomous System incorporates. Once the Maxmind (2015) Geo IP2 database was downloaded, we opened the 'complete_IP_import_for_gephi.xlsx' file and created a new sheet in the above-mentioned file named 'MaxMindGeoIP2'. We imported the downloaded *.csv file of the Maxmind (2015) Geo IP2 database, which provided us with the served IP address ranges for each Autonomous System in a four-octet format. For the purposes of transforming the IP addresses to their Autonomous System Numbers, we first determined the dotted String values (e.g. '217'. '225'. '240'. '18') of all IP addresses (including destination IP address, source IP address and hop IP address) and transformed them into their respective four-

octet values (O_1, O_2, O_3 , and O_4). Next, we calculated as follows:

$$O_1 * 256^3 + O_2 * 256^2 + O_3 * 256 + O_4$$

Once the IP addresses were transformed to their four-octet values, we saved them in a new column. Next, we fused the Maxmind (2015) GeoIP2 database from the second sheet to the first sheet in using Excel's VLOOKUP function, which searched for each transformed four-octet IP address value for the corresponding Autonomous System Number and Autonomous System Name in the 'MaxMindGeoIP2' sheet. The corresponding file was then saved as 'IP_to_ASN.xlsx', replacing the 'complete_IP_import_for_gephi.xlsx' one. To obtain the Autonomous System Numbers for each IP address of the three mobile broadband operators, we simply filtered the generated dataset by using our campaign identifiers.

After the transformation from IP address to AS Number, we replaced the destination IP address, source IP address and hop IP address columns in the 'Traceroute_Raw_Data_Analysis.xlsx' file by the corresponding mapped Autonomous System Numbers and changed the headers to 'ASNdestination', 'sourceASN' and 'hopASN' and saved the file. We then separated the *traceroute* hop observations of interest using the campaign identifier as described above. Next, we copied the 'sourceASN' and 'hopAS' into new sheets and added their 'Source' and 'Target' header rows, respectively. The sheets were then saved as *.xlsx and *.csv files again (e.g. as 'ASN_55831_for_gephi.xlsx' and 'ASN_55831_for_gephi.csv' for Aircel). Next, we imported the *.csv files as directed edges tables, one after another, into Gephi (2016) and calculated the relevant Complex Network metrics using the Gephi (2016) Statistics settings. After saving the elaborated file (e.g. as 'AS45609_after_gephi.gephi' for Bharti Airtel), we reported our obtained findings in section 4.3 in Chapter 4.

3.5.5 Iteration 5: Graph Visualisation Analysis (AS)

Just like Iteration 3 at IP granularity, this Iteration 5 at Autonomous System granularity aimed, based on the obtained understanding from Iteration 4, to approve the previous findings that the three Tamil Nadu mobile broadband operator networks really followed *Scale-Free Network* properties. Moreover, similar to Iteration 3, we aimed to generate insight into the connectivity importance of certain vertices that might indicate structural bottlenecks (given the apparent *Scale-Free Network* nature) for providing upstream internetworking features of the three mobile broadband operator networks, here at Autonomous System granularity. In addition to Iteration 3, we also generated graph visualisations of the different Centrality metrics (as introduced in section 3.4.2 above). Doing so also aimed to gain a deeper understanding of the three mobile broadband operator networks. Hence, this section starts by describing the steps to generate the *Small-World Network* Model and the Barabási-Albert *Scale-Free Network* Models followed by the Centrality Metrics and lastly the *k-core decomposition*. The outcomes of Iteration 5 were reported in the section 4.3.7 of Chapter 4.

Small-World and Scale-Free Models

The *Graph Visualisation Analysis* at Autonomous System granularity followed the same approach to the previous Iteration 3 at IP granularity, including the layout parameters, (see section 3.5.3) while utilising the AS mapped *.csv files from Iteration 4. Here, the Gephi (2016) files for the *Small-World Network* Model were saved as ‘AS55831_Kleinberg_blue.gephi’, while the corresponding graph visualisation *.png files were saved as ‘AS55831_Kleinberg_blue.png’ for Aircel in the respective sub-folder. Similarly, the *Scale-Free Network* Model files were generated in using Gephi (2016) and saved as ‘AS38266_BAModel_blue.gephi’ and ‘AS38266_BAModel_blue.png’ for the Barabási-Albert Standard Model.

Centrality Metrics

In addition to Iteration 3, the *Graph Visualisation Analysis* at Iteration 5 (AS granularity) generated further Centrality metrics. For this purpose, we again imported the mapped *.csv files from Iteration 4 as directed edge-table into Gephi (2016) for each of the four utilised Centrality metrics (*Degree Centrality*, *Closeness Centrality*, *Betweenness Centrality* and *Eigenvector Centrality*). Hence, the next step after the import covered the consecutive calculation of the respective Centrality metrics. Once calculated, we set the graph visualisation layout to the force-directed *Fruchterman-Reingold* (1991) *Layout*.

We then set the vertex colouring ‘red’ for those vertices with low centrality values and ‘blue’ for those with high centrality values. This is achieved by first selecting the corresponding Centrality metric by following ‘Appearance > Nodes > Attributes’. Next, we set the respective colouring by using the colour slider and applied the settings. Next, we provided the graph visualisation with the layout properties as mentioned in Chapter 4, saved the Gephi (2016) files as ‘AS38266_Eigenvector_Centrality.gephi’ for example for the *Eigenvector Centrality* of Vodafone and exported the corresponding graph visualisations as *.png files, here ‘AS38266_Eigenvector_Centrality.png’. Lastly, we reported the graph visualisations alongside the findings for each mobile broadband operator in section 4.3.7.

k-core decomposition

The *k-core decomposition* at AS granularity followed the same approach to the previous Iteration 3 at IP granularity, while again utilising the AS mapped *.csv files from Iteration 4. This file was again cleaned from the header rows and identifier columns and saved as ‘AS38266_Vodafone_for_R.csv’ for Vodafone, before importing it in the Statistical Computing Tool R (2016). Here, the Network Analysis and Visualisation package ‘igraph’ was now already installed, meaning we could straightaway adjust the R-script from Iteration 3 for the one at Iteration 5, replacing the utilised file as well as some of the graph visualisation settings. After computing the elaborated R-script in R (2016), we saved the resulting graph visualisations (e.g. as ‘AS38266_Vodafone_k_core_decomposition.png’ for Vodafone) and reported our findings from the R (2016) console in Chapter 4.

3.5.6 Iteration 6: Autonomous System Relationships

The *k-core decomposition* of the *Graph Visualisation Analysis* in Iteration 5 indicated an important set of densely connecting Autonomous Systems for each of the three Tamil Nadu mobile broadband operators. Merely identifying these structural bottlenecks was not satisfactory. Hence, Iteration 6 aimed to reveal the economic nature of the most important mobile broadband operator networks’ relationships between the influential Autonomous Systems. The outcomes of Iteration 6 are reported in section 4.4 of Chapter 4.

Secondary Datasets

To reveal the economic nature of the Autonomous System relationships, we first opened the CAIDA (2016a) AS-Rank website and filtered the visualised AS-Rank dataset to fit the data-collection time of this dissertation's time horizon (see section 3.3.4). The filtering was therefore set as 'Dataset: 2015-02-01 IPv4'. Once the table view updated, it revealed 49,874 Autonomous Systems containing information on the Customer Cone Size (Number of Autonomous Systems and IPv4 prefixes), the percentage of the AS of all Autonomous Systems, IPv4 prefixes and the AS Transit Degree. Next, we sorted the table view by 'number of ASes in customer cone', resulting in another reload of the table view. We then downloaded the *.html table contents and placed them in an Excel file with the following steps:

- Right-click on the CAIDA (2016a) AS-Rank website > view page source
- Copying all the content in the cache.
- Pasting the copied content in a new text document with 'Ctrl+v'
- Saving the file as 'as-rank.html',
- Converting the 'as-rank.html' file into a *.csv file by using Conversiontools (2012).
- Saving the file as 'CAIDA_AS_Rank_Data_01-02-2015.csv' and 'CAIDA_AS_Rank_Data_01-02-2015.xlsx'.

Once downloaded, we manually searched the resulting file according to the most important information (Customer Cone Size, Number of IPv4 prefixes and Transit Degree) for those Autonomous Systems that the *k-core decomposition* in Iteration 5 revealed as most interesting to our case study. Additionally, we referred to the Border Gateway Protocol (BGP) Routing Tables of Hurricane Electric (2016). This helped us to provide a more thorough understanding of our operator networks, where applicable. Some of these tables are stated in the Appendices. Next, we downloaded the secondary CAIDA(2016b) AS-Relationship dataset by filling out the prompted user info request on the CAIDA website. This secondary dataset helped us to test our three Tamil Nadu mobile broadband operator networks through the economic nature of Autonomous System relationships, where a relationship could either be of *peer-to-peer*, *customer-to-provider*, or *provider-to-customer* nature. The downloaded file was saved as 'CAIDA_AS_Relationship_Data.txt'.

Once obtained, we opened the operator's Autonomous System edge tables that resulted through the Gephi (2016) export (containing the source and hop Autonomous System Numbers) in Iteration 4 and saved the file as 'Aircel_ASRank_ASRel_mapping.xlsx' and 'Aircel_ASRank_ASRel_mapping.csv', respectively. The data itself is stored in a sheet named e.g. 'Bharti_Airtel_edges_after_Gephi' for Bharti Airtel. We then added two new sheets to the file, named 'AS_rel' and 'Transit_Table'. The first sheet, 'AS_rel', contained the AS-Relationship data from the 'CAIDA_AS_Relationship_Data.txt', filtered by those Autonomous Systems of the respective mobile broadband operators. This data resulted from the secondary CAIDA (2016b) AS-Relationship dataset. The second sheet, 'Transit_Table', contained the Transit Degree obtained from the 'CAIDA_AS_Rank_Data_01-02-2015.xlsx' file from the secondary CAIDA (2016a) AS-Rank dataset.

Next, using Excel's INDEX algorithm (see algorithm in Excel file), we fused the source and hop Autonomous System Numbers in the first sheet (e.g. 'Bharti_Airtel_edges_after_Gephi') with their associated Transit Degree from the elaborated 'Transit_Table' sheets, into two new columns in the first sheet, named 'Source_Transit' and 'Target_Transit'. We then fused the source and hop Autonomous System Numbers, together with their corresponding 'Source_Transit' and 'Target_Transit', which resulted in two new columns containing the 'SourceASN: TransitDegree' and the 'HopASN: TransitDegree', respectively. To prepare for the later Gephi (2016) import, we named these new columns 'Source' and 'Target' and saved the file as 'AS_Rank_Analysis.xlsx'.

Since the goal of this analysis is to explore the economic relationships of the Autonomous Systems in the operator networks, we next fused the file with the secondary CAIDA (2016b) AS-Relationships dataset. For this purpose, we first imported the downloaded 'CAIDA_AS_Relationship_Data.txt' into Excel, saved the sheet as 'AS_rel' and the file as 'AS_Rel_Analysis.xlsx' in the './Step6/Secondary_CAIDA(2016b)_AS_Relationship/' folder. The AS-Relationships in the 'AS_rel' sheet are represented by three columns, named 'AS1', 'AS2' and 'rel', indicating the relationship between two Autonomous Systems. Next, we created three sheets named 'Aircel', 'Bharti Airtel' and 'Vodafone' and imported the respective operator edge-tables from Iteration 4. Each of these three sheets contained only a 'Source' and a 'Target' column. To uniquely match the relationships for the three operator networks, we first combined the 'AS1' and 'AS2'

columns in the ‘AS_rel’ sheet into a new column, named ‘AS1AS2’. This column was useful for unique referencing purposes. Similarly, we fused the ‘Source’ and ‘Target’ columns for each of the three mobile broadband operator sheets (see above) into a column named ‘SourceTarget’ column. We then added a new column named ‘INDEXMATCH’ to each operator sheet, where we used a combination of Excel’s MATCH and INDEX algorithms, linking the ‘AS1AS2’ column of the ‘AS_rel’ sheet with the ‘SourceTarget’ columns of the operator sheets to reveal the corresponding ‘rel’ column of the ‘AS_rel’ sheet for each of the given operator sheets. The result represented the AS-Relationships (*‘peer-to-peer’*, *‘customer-to-provider’* and *‘provider-to-customer’*) for the Autonomous Systems in our three operator networks, represented by a ‘0’ for a *peer-to-peer* relationship, a ‘-1’ for a *provider-to-customer* relationship and a ‘1’ for a *customer-to-provider* relationship.

Next, looking at a combination of both, the ‘AS_Rel_Analysis.xlsx’ and the ‘AS_Rank_Analysis.xlsx’, we reported some preliminary findings in section 4.4 for each of the three mobile broadband operators. Next, we created a new file for each of our case studies’ three operator networks named e.g. ‘Aircel_ASRank_ASRel_mapping_(Edges).xlsx’ for Aircel. In this file, we copied the ‘Source_Transit’ and ‘Target_Transit’ columns from the ‘AS_Rank_Analysis.xlsx’ file as well as the corresponding ‘rel’ column from the ‘AS_Rel_Analysis.xlsx’. This allowed us to measure the economic relationships per mobile broadband operator. For this purpose, we created a sheet named ‘analysis’ for each of the three operator files and counted the number of edge observations, the number of edge-weights and the percentage of edge weights of all edges per AS-Relationships, as stated above. Our findings were again reported for each mobile broadband operator. Moreover, to visualise the economic relationships between the Autonomous Systems in the three operator networks, we visualised the three networks again in a two-dimensional Euclidean space. For this purpose, we generated a *.csv file (named e.g. ‘Aircel_ASRank_ASRel_mapping_(Edges).csv’ for Aircel) from the respective Excel files (named ‘Aircel_ASRank_ASRel__mapping_(Edges).xlsx’ for Aircel, for example) and imported the *.csv files again as edge table into Gephi (2016) and saved them (named ‘Vodafone_ASRank_ASRel_mapping.gephi’ for Vodafone, for example). The key here was to set the ‘rel’ column as relationship label when importing the dataset into Gephi (2016). This allowed us to colour the relationships, or edges, between each set of Autonomous Systems. Here, we coloured a *peer-to-peer* relationship between a set of

Autonomous Systems ‘green’, *provider-to-customer* relationships ‘red’, ‘*customer-to-provider*’ ones ‘blue’ and ‘yellow’ for undetected ones. The resulting graph visualisations were then saved, using weighted and non-weighted edges, as *.png files (named, for example, as ‘Aircel_Relationships_w.png’ for the Aircel graph visualisation). Next, we again reported our findings, for each of the three operator networks.

3.5.7 Iteration 7: Statistical Network Analysis

Of our findings from the previous Iterations 1-6 above, Iteration 7 aimed to provide more confidence towards the applicability of the exploratorily derived indications. Moreover, we aimed to link the apparent hierarchical upstream Internet market structuring of the three Tamil Nadu mobile broadband operators. We aimed to reveal the potential effects on the affordability of the respective mobile broadband operator price plans, as measured in price per Megabyte. This was a crucial step to provide further explanatory evidence and applicability of our findings. The following sections are separated given the employed two-stage process (described in Chapter 5). The section below covers the data preparation steps for two econometric in the first stage of the process, namely Model 1 and Model 2, followed by the econometric models of stage 2, namely Model 3.1, Model 3.2 and Model 4. Further information and outcomes of this last iteration were reported in the sections 5.3 and 5.4 in Chapter 5.

Model 1 and Model 2

The first two Econometric Models connected the network structural markers of the operator networks to the derived metrics for the upstream connectivity. First we prepared the data for the anticipated analysis. Here, we exported the vertex tables as *.csv files for each operator network at Autonomous System granularity from Gephi (2016) by using the Gephi (2016) operator network files obtained in Step 4. Next, we generated a new Excel file named ‘Model1_for_Stata.xlsx’. Here, we pasted the exported operator *.csv files into new sheets called ‘Aircel’, ‘Bharti Airtel’, and ‘Vodafone’. We then generated a new sheet named ‘merged’, where we successively pasted the data from the three provider sheets while providing them with a new column named ‘prov’. This column represents the operator of interest, where ‘1’ refers to Aircel, ‘2’ for Bharti Airtel and ‘3’ for Vodafone. This column allowed us to filter the statistics per operator later. We then opened the Data Analysis and Statistics Software Stata (2016), imported the ‘Model1_for_Stata.xlsx’ file and saved the resulting Stata (2016) data file as ‘Model1_2_data.dta’. The elaboration of the econometric models in Chapter 5 were then

stored in a Stata do-file called ‘Model1_2.do’. The script of this Stata do-file, including their line-by-line description, is enclosed in the Appendices, while the results of the Model 1 and Model 2 were reported in section 5.3 and discussed in section 6.1.

Model 3.1 and Model 3.2

The third econometric model in the second stage links the coefficients obtained in Model 1 and Model 2 (capturing the relationship between network structural markers of the operator networks with their *Weighted Out-Degree* and *Weighted In-Degree* upstream connectivity) to the price plans of our three mobile broadband operators. In detail, we generated a new Excel file, named ‘Model3.csv’. Next, we stored the price plan information (columns: ‘datainmb’ for data allowance in Megabyte, ‘vin’ for validity in days, ‘Price’ and ‘pricepermb’ for price per Megabyte) obtained from GSMOutlook (2015a, 2015b, 2015c) and marked each observation with an identifier in the ‘id’ column. Additionally, we added the coefficients obtained from Model 1 into the columns labelled ‘lclus_hat’ and ‘leige_hat’ and marked them with their corresponding mobile broadband operator in the ‘prov’ column. Similar to the previous steps, ‘1’ corresponded to the in Model 1 estimated coefficients of Aircel, ‘2’ for the coefficients of Bharti Airtel and ‘3’ for the Vodafone ones. We then opened Stata (2016) and imported the generated ‘Model3.csv’ file. The specification of the econometric model was then stored in a Stata do-file called ‘model3.do’, containing Model 3.1 and Model 3.2. The script of this Stata do-file is enclosed in the Appendices, while the results of the Model 3.1 and Model 3.2 were reported in section 5.4 and discussed in section 6.2.

Model 4

The fourth econometric Model in the second stage associates the coefficients obtained in Model 2 (the relationship between network structural markers of the operator networks with their *Weighted In-Degree* upstream connectivity) with the price plans of our three mobile broadband operators. For that purpose, we copied the Model 3 generated ‘Model3.csv’ file and replaced the values of the ‘lclus_hat’ and ‘leige_hat’ columns with the coefficients that we obtained in Model 2. The price plan observations remained unchanged. Next, and in a similar way to Model 3, we then opened Stata (2016) and imported the changed ‘Model4.csv’ file. The steps of the econometric model’s elaboration were then stored in a Stata do-file named ‘Model4.do’. The script of this Stata do-file, including their detailed description was again enclosed in the Appendices, while the results were reported and discussed in section 5.4.

Correlation Table

The correlation table associates the coefficients obtained from Model 1 and Model 2 with Quality of Service data of TRAI. The do-file was enclosed in the Appendices.

3.6 Ethics, Biases, Reliability, Validity & Generalisability

This section states the potential biases that the work in this dissertation might have confronted. Additionally, we attached our statement on Ethical considerations, reliability and generalisability in the Appendices.

3.6.1 Biases

Here, we critically examine any potential research biases with which our case study may have been confronted. Alongside the research process, we took great caution to avoid any presence, behaviour or attitudes affecting the *traceroute* data collection, their inherent measurements or the following exploratory Network Analysis. Hence, this section includes our statements on Selection Bias, Inclusive Bias, Measurement Bias and Reporting Bias.

Sample Selection Bias

In any form of research, it would be ideal, but inherently too costly, time-consuming and often impractical, to study the entire population. The *traceroute* data collection presented in this dissertation followed a mobile crowdsourcing approach from an active Internet periphery perspective where three Android smartphones were used for the primary data collection. Even though our single case study was chosen strategically, the three Android smartphones and their SIM cards may still have unconsciously been selected for convenience purposes. This choice included i) easy access to the chosen devices and ii) easy access to the chosen SIM cards. Hence, we assume a sample selection bias due to convenience sampling, where the results can neither be extrapolated to other smartphone producers, nor for the entire population of mobile broadband operators (SIM cards) in Tamil Nadu, India, or elsewhere. An ideal sample would have included a local population of available smartphones as well as the full amount of available SIM cards from all given local providers in Tamil Nadu, India. The mitigation of this sample selection bias could not have been overcome using statistical analysis. Scott and Carrington (2011) refer to one lone example where the sample selection bias for Network Analysis was mitigated using a *Heckman Selection* model. Despite this indication, we were not utilising the *Heckman Selection* model. Hence, we accounted for this sample selection bias when

reporting our analytical results.

Measurement Bias

Measurement bias is a result of poorly measuring the case study object of interest. Since we chose to follow a mobile crowdsourcing *traceroute* data collection technique, using the Portolan (2015), we delineate ourselves from any form of measurement biases. At times during the data collection period (measurement campaign), the Portolan (2015) Android application was not revealing any measurements. However, these data collection failures represent vital measurement steps that are not associated with poorly measuring the object of interest, but instead thoroughly measuring the actual mobile broadband operator situation in Tamil Nadu, potentially employing poor mobile broadband coverages or certain smartphone issues.

Reporting Bias

The Reporting bias refers to the underreporting of unexpected or undesirable results in this case study. In the course of our abductive inquiry process, it was crucial to gain an understanding of the nature of the three Tamil Nadu mobile broadband operator networks through our collected *traceroute* measurements. Hence, we were cautiously reporting all analytical results equally, independent of their given characteristics, properties or values. Hence, we assume that our research does encounter very minor reporting biases, although the probability of under-reporting findings and evidence may still exist, since the collected *traceroute* data covers a wide complexity and applicability. Furthermore, and given the due course of our *Working Hypotheses*, we were not able to study all possible Network Analysis metrics but covered those that aimed to reveal the greatest understanding for our given research problem.

3.7 Summary

This Chapter started by laying out our philosophical assumptions that determined the scope and limitations of this dissertation. We justified our choice of a pragmatist paradigm to research philosophy, followed by an explanation of our abductive approach to research. Based on these underlying assumptions, we then reasoned our strategic choice for a single case study strategy that incorporates an exploratory-quantitative multimethod design based on *Descriptive*, *Complex Network*, *Graph Visualisation*, and *Statistical Network Analysis* methods. All of these previous steps informed our choice of time horizon followed by our crowdsourced primary *traceroute* data collection technique

using active Internet periphery measurements. We closed this chapter by stating the employed analytical procedures as inquiry iterations in detail. Summarising, this chapter aimed to help the reader to replicate our case study findings. It also provided a methodological frameset for use in a future analysis of similarly strategically relevant cases that aim to advance the field of studying the upstream Internet structures and the importance of structural bottlenecks in developing and low-middle income countries.

4 COMPLEX NETWORK ANALYSIS

‘The Web is now philosophical engineering. Physics and the Web are both about the relationships between the small and the large’ (Berners-Lee, 2015).

The Internet as a network of networks represents a complex system of interacting social, economic and technical infrastructures. Its sheer complexity makes it a non-trivial task to gain a holistic understanding of structural phenomena, the single importance of economic relationships between interacting agents such as *Internet Service Providers* and the underlying economic relationships of these relationships for creating global internetworking connectivity. Network Sciences provides us with helpful mathematics-, physics-, and computer science-based methods for exploring these structural relationships between interacting agents. Computer Sciences, in particular, grant us deeper insights into how communication networks might be analysed. Our aim is to understand the operator networks and their structuring as a whole as well as the importance of distinct and influential agents as bottleneck parts that compose these networks.

Therefore, this chapter starts by describing the obtained data from the active Internet periphery measurements before exploring them in their natural state at Internet Protocol granularity. This granularity represents the network reality in the form of connections amongst machines, identified through their unique IP addresses. Here, we aim to reveal the general structural properties of the operator networks. Next, this chapter associates the IP addresses with their operating entities, called Autonomous Systems. Through this granularity, we aim to understand the changes in structural properties from Internet Protocol to Autonomous System granularity as well as the general structural organisation in the operator networks, revealing *Autonomous Systems* with key structural properties. Moreover, we aim to expose distinct internetworking relationships between the major *Internet Service Providers*. This is a crucial precondition for understanding the underlying economics governing our three operator networks. Next, linking the operator networks with a secondary dataset, we aim to expose the economic nature of the most relevant relationships amongst Autonomous Systems. Lastly, we summarise the findings of this chapter to provide a holistic view on the Tamil Nadu mobile broadband operator networks.

4.1 Descriptive Network Analysis

The exploration commences by analysing the essential features of the *traceroute* data collected using our active Internet periphery measurements through the Portolan (2015) (see section 3.3.4). Figure 4-1 below provides a look and feel for our collected 731,200 individual *traceroute* hop observations.

Traceroute ID	Source IP	Destination IP	AS Path	Hop Number	RTT (ms)	Loss (%)
112165856	91.198.69.9	91.198.69.9	WORLDin24560background	1	480.773	0
112165856	103.14.161.1	103.14.161.1	WORLDin24560background	2	0	0
112165856	91.223.229.130	91.223.229.130	WORLDin24560background	3	32.6811	0
112165858	91.223.229.130	91.223.229.130	WORLDin24560background	1	31.3161	0
112165858	125.21.0.105	125.21.0.105	WORLDin24560background	2	259.471	0
112165858	125.62.187.113	125.62.187.113	WORLDin24560background	3	0	0
112165859	103.245.22.88	103.245.22.88	WORLDin24560background	1	32.6896	0
112165859	103.245.22.88	103.245.22.88	WORLDin24560background	2	31.4929	0
112165859	103.245.22.88	103.245.22.88	WORLDin24560background	3	91.8489	0
112165859	103.245.22.88	103.245.22.88	WORLDin24560background	4	85.5242	0
112165859	103.245.22.88	103.245.22.88	WORLDin24560background	5	92.4238	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	1	32.3535	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	2	32.2652	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	3	101.388	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	4	94.3622	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	5	101.845	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	6	95.3473	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	7	102.383	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	8	114.993	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	9	109.978	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	10	115.564	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	11	116.549	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	12	109.929	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	13	0	0
112165860	122.200.3.3	122.200.3.3	WORLDin24560background	14	32.4023	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	1	30.8787	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	2	0	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	3	0	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	4	31.1712	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	5	31.9038	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	6	0	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	7	32.7072	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	8	31.4056	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	9	264.028	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	10	294.142	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	11	294.973	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	12	325.079	0
112165861	186.251.208.25	186.251.208.25	WORLDin24560background	13	327.794	0

Figure 4-1: Example of collected Paris traceroute observations.

As the different columns in Figure 4-1 above indicate, each *traceroute* hop (consist of a source IP address linking to a destination IP address) observation contains the following information:

- *Traceroute* identifier.
- The randomly-chosen destination of a given *traceroute*.
- Campaign identifier consisting of an identifier for the associated country of initial connection (‘WORLDin’ indicates India) and an identifier for the Autonomous System Number (e.g. ‘24560’) of the initial connection.
- Timestamp, comprised of YYYY-MM-DD and the exact record time.
- Geo-location (Latitude, Longitude) of the data-collecting device.
- The operating system of the data-collecting device (e.g. ‘android’).
- Associated hop number of a *traceroute* (e.g. third hop / step of a given *traceroute*) to which the row refers (primary observation unit used in the following Complex and Statistical Network Analysis in the rest of the dissertation).
- Source IP address (starting point of the hop within a *traceroute*).

- Target IP address (arrival point of the hop within a *traceroute*).
- Round Trip Time (RTT) of a given hop.
- Binary indication whether or not a *traceroute* hop observation contains a skip (e.g. ‘1’ representing a failure when the connection between a source IP address and a destination IP address in the hop is not reachable or terminates, ‘0’ otherwise).

The complete dataset indicated the presence of *traceroutes* starting from both mobile broadband and Wi-Fi connections covering different locations. Here, only those measurements of the three Tamil Nadu mobile broadband operators were of interest. Hence, we separated the *traceroute* observations originating from Wi-Fi, from those *traceroutes* originating from the mobile broadband operators that will be used to compare our three operators of interest thoroughly. Given the nature of a *traceroute*, each collected observation incorporates a multitude of *traceroute* hops (or steps along a connection). Filtering the data by the identifier revealed that the total *traceroute* hop observations consisted of 57,121 unique *traceroutes* (each containing multiple hops), including those originating from Wi-Fi connections. The randomly-chosen destinations further exposed that the Portolan (2015) application randomly assigned 32,068 unique destinations for these 57,121 *traceroute* observations. Here, the random selection is used to replicate, in a possibly unbiased way, the behavioural patterns of end-users. Moreover, the campaign identifiers revealed that the recorded *traceroute* hop observations were commencing from twelve distinct Autonomous Systems Numbers (see Table 9-9 in the Appendices). By using the Hurricane Electric (2016) BGP-Toolkit, these campaign identifiers were associated with their organisational name. Revealing these names allowed us to choose only those non-Wi-Fi originating observations that are of fundamental interest for our analysis of the mobile operators’ upstream connectivity. Hence, this step was crucial in selecting and filtering the relevant dataset that will be used in the following analysis. We implemented this step by verifying the campaign identifiers using the Maxmind (2015) GeoIP2 database, together with UltraTools (2016), Team Cymru (2016) and the Hurricane Electric (2016) BGP Toolkit. Linking the Autonomous System Numbers to the collected *traceroute* hop dataset revealed that most of these collected hop observations belonged to *traceroutes* originating from the Wi-Fi based Spectranet (AS10029) and C48 Okhla Industrial Estate (AS55410), see Appendices.

After filtering out the set of observations originating from Wi-Fi connections we were left with only 36,388 total *traceroute* hop observations being relevant to our case study. Those represent the only connections originating from the Tamil Nadu mobile broadband operators. More specifically, Vodafone indicated 30,633 mobile broadband observations, followed by Aircel with 4,749 ones and Bharti Airtel with 956 observations. The filtered out 649,812 *traceroute* hop observations resulted, as the time stamps confirm, from Wi-Fi connections mainly captured during off and night-time hours by the Android smartphones, also containing observations from other locations.

Table 4-1 below indicates the total number of *traceroutes* (column 2 in Table 4-1) and the total number of *traceroute* hop observations (column 3 in Table 4-1) that were obtained for each of the three Tamil Nadu mobile broadband operators. Here, the average number of hops per *traceroute* observation is interesting since it indicates that Vodafone needed, on average, considerably more hops to complete a given *traceroute* than the other two operators.

Traceroute hop observations by operator			
Mobile broadband operator	Number of <i>traceroute</i> observations per mobile broadband operator	Number of <i>traceroute</i> hop observations contained in all the mobile broadband operator-originated traceroutes	Average Number of <i>traceroute</i> hop observations per <i>traceroute</i> observation
Aircel	622	4,749	7.63
Bharti Airtel	148	956	6.46
Vodafone	2,678	30,633	11.44

Table 4-1: Traceroute hop observations by mobile broadband operator.

The associated number of hops describes the actual number of steps that a *traceroute* needed to take in order to reach its randomly-assigned final destination, through the routers, identified via their unique IP addresses, forming the basic steps of the observed internetworking through the Internet. Interestingly, we discovered that the path length of the Aircel *traceroute* observations ranged between 5-40 hops, the Bharti Airtel one ranged between 4-36, and the Vodafone one between 5-51 hops.

The Round-Trip-Time (RTT) indicates the time a data packet takes to be sent from the

initial source IP address to the destination one, plus the time it takes for this to be acknowledged by the destination IP address and returned to the source IP address. An analysis of the RTT indicated, as the following Table 4-2 illustrates, that the lowest Round-Trip-Time of a completed *traceroute* was reached by Vodafone (0,042ms). These results are, in contradiction with the previous ones on the average hops per *traceroute* observation of each mobile broadband operator. This might indicate that there is a large amount of potential connections between IP addresses belonging to the same Autonomous System, as we will explain later in more detail. Furthermore, the Bharti Airtel observations revealed the largest range between the lowest and highest Round-Trip-Times, indicating that end-users might experience fluctuations in their perceived Quality of Service (QoS).

Round-Trip-Time (RTT) by mobile broadband operator			
Mobile broadband operator	RTT Low in ms	RTT High in ms	Range
Aircel	0,042384	1006,07	1006.028
Bharti Airtel	0,106708	2019,58	2019.473
Vodafone	0,044219	1243,9	1243.856
Key ms: milliseconds. RTT: Round-Trip-Time.			

Table 4-2: Round-Trip-Time by mobile broadband operator.

Table 4-3 below displays the skip-distributions for each of the three operators, i.e. the frequency distribution of *traceroutes*, depending on the specific step (hop) along the *traceroute*, where the connection fails (skips). Comparing the mobile broadband operator *traceroute* skip-distributions potentially revealed another indicator for perceived Quality of Service.

Skip distribution by mobile broadband operator										
Mobile broadband operator	Skip 0	Skip 1	Skip 2	Skip 3	Skip 4	Skip 5	Skip 6	Skip 7	Skip 8	Skip 9
Vodafone	30,240	197	96	36	26	11	6	7	6	8
Bharti Airtel	884	58	13	1	0	0	0	0	0	0
Aircel	4,670	60	11	3	4	1	0	0	0	0

Table 4-3: Skip distribution by mobile broadband operator.

Summarising, the above *Descriptive Network Analysis* provides initial insights about some aspects of the internetworking features of the three mobile broadband operators, eventually affecting Quality of Service from an Internet periphery perspective. We showed that 36,388 of the 731,200 *traceroute* hop observations were relevant for this case study and compared the general properties of the observations per mobile broadband operator of interest. The following section aims to further uncover the distinct connectivity features in the upstream Internet access market of the three mobile broadband operator networks using *Complex Network Analysis*.

4.2 Complex Network Analysis at IP granularity

This section focuses on the three mobile broadband operator networks at Internet Protocol granularity. This granularity represents connections between IP addresses, indicating the unique identifiers of machines on the Internet. We start this section with an exploratory analysis of the general network metrics per mobile broadband operator. Next, we study the relevant edge and vertex metrics, followed by a *Graph Visualisation Analysis* before concluding on the Complex Network properties of these networks.

4.2.1 General Network Metric Analysis (IP)

General network metrics are used to describe some key structural properties of the network generated by the set of internetworking connections originating from each of the three mobile broadband operators. After describing the set of *traceroute* hop observations, we look at the *Average (Weighted) Degree*, the *Network Diameter*, the *Network Density* and the randomized *Modularity* for each of the three generated networks.

By utilising the ‘source IP address’ and the ‘target IP address’ of the identified 36,388

traceroute hop observations (after filtering out the Wi-Fi originating ones), the three mobile broadband operator networks may each be denoted as a directed graph, $G_{Operator} = G(V, E)$. Here, $V_{Operator}$ represents the total number of IPv4 address vertices being traversed by the data packets along the entire set of *traceroutes*, originating from the specific mobile broadband operator. While $E_{Operator}$ represents the total number of edges connecting distinct pairs of the above vertices $V_{Operator}$, where $E_{Operator} = \{\{u, v\} | u, v \in V_{Operator}\}$. Hence, three resulting directed networks, generated for the three mobile broadband operators, are denoted as: G_{Aircel} , $G_{Bharti Airtel}$ and $G_{Vodafone}$. Each of these three networks represents a sub-network of the overall network $G_{Operator}$, comprising the entire set of *traceroute* hop observations at IP granularity from each mobile broadband operator, $G_{Aircel}, G_{Bharti Airtel}, G_{Vodafone} \in G_{Operator}$.

Exploring the general metrics for these three directed networks, G_{Aircel} , $G_{Bharti Airtel}$ and $G_{Vodafone}$ shows that G_{Aircel} consisted of 2,259 unique IP address vertices and 2,879 edges (with repetition) connecting those vertices. Therefore, we denote this network as $G_{Aircel} (2259, 2879)$. Hence, the other two directed networks are denoted as $G_{Bharti Airtel} (600, 803)$ and $G_{Vodafone} (7509, 10390)$, respectively. All edges stated above included repetitions, which means that many connections between a pair of vertices are potentially traversed multiple times. These often-used edges are crucial when calculating the edge weightings in a given network graph, representing recurring internetworking connections at IP granularity. An important note is that the IPv4 address '0.0.0.0' was consistently reached once a *traceroute* terminated. This IP address is, therefore, expected to show a large number of incoming connections, assuming that the *traceroutes* are reaching their destinations.

The first metric we analyse is the *Average Degree*, given by the ratio of the total number of edges $E_{Operator}$ over the total number of vertices $V_{Operator}$, $\frac{E}{V}$. This metric provides a first indication on the edge density in the mobile broadband operators' networks. This metric revealed minor density differences among the operators. Here, G_{Aircel} indicated 1.274 edges per vertex, followed by $G_{Bharti Airtel}$ with 1.338 and $G_{Vodafone}$ with 1.384 ones. These differences indicate that Aircel required, on average, fewer connections per IP address vertex than the other two networks, indicating the regular usage of certain IP address vertices.

The *Average Weighted Degree*, as being the average of an IP address vertex connectivity

in a given operator network, provides further interesting differences among the three operators. While $G_{Bharti\ Airtel}$ showed the smallest *Average Weighted Degree* of 1.593, G_{Aircel} indicated an *Average Weighted Degree* of 2.102 and $G_{Vodafone}$ showed the highest value of 4.080 average number of connections for IP address vertices belonging to $G_{Vodafone}$.

Next, we report the values for the mobile broadband operator graphs' *Network Diameter* (directed), a metric that indicates the longest possible shortest distance paths between any two IP address vertices (see section 3.4.2). The implication of a longer *Network Diameter* could affect the Quality of Service of a given mobile broadband operator as this negatively affects longer path lengths.

The overall small values of the mobile broadband operators' Graph Density (directed), see Table 4-4 below, indicate that IP address vertices are not densely connecting between each other. This would come as no surprise given the nature of a *traceroute*. However, we expect that a small number of IP addresses are more densely connected.

The *Randomised Modularity* is an additional metric capturing the structure and dynamics of a network (Newman, 2006). Our data revealed the existence of 168 vertex clusters in G_{Aircel} , 147 in $G_{Vodafone}$ and only 61 in $G_{Bharti\ Airtel}$. This indicates that G_{Aircel} has the densest connections between vertices in clusters but sparse connections between vertices linking those clusters, representing stronger structuring. The *Modularity* metric itself is stated in Table 4-4 below.

General Metrics by mobile broadband operator at IP granularity					
Mobile broadband operator graph	Average Degree	Average Weighted Degree	Diameter (Directed)	Density (Directed)	Randomised Modularity
G_{Aircel}	1.274	2.102	17	.001	.691
$G_{Bharti Airtel}$	1.338	1.593	14	.002	.627
$G_{Vodafone}$	1.384	4.080	25	.000	.669
Key IP: Internet Protocol.					

Table 4-4: General metrics by mobile broadband operator at IP granularity.

In the next section, we will analyse additional edge metrics with the objective of uncovering more structural features for our three mobile broadband operator networks.

4.2.2 Edge Metric Analysis (IP)

The edge metrics discussed in this section focus on the connections between any given set of vertices in an operator network. We analyse the *Average Path Length*, the *Average Neighbourhood Overlap* and the *Average Embeddedness* edge metrics.

First, the Average Path Length (see Chapter 3) was, with a value of 6.404, considerably higher for $G_{Vodafone}$ than for the other two networks. Table 4-5 below illustrates these findings. These differences might indicate two distinct situations. First, that Bharti Airtel routes connections more efficiently, or second that Vodafone and Aircel more heavily rely on AS-internal routings.

The second analysed edge metric was the *Average Neighbourhood Overlap*. This edge metric represents the intersection of the union between neighbouring vertices in a given network (see edge metrics in section 3.4.2). Moreover, the findings for the *Average Embeddedness* support those of the *Average Neighbourhood Overlap*. This edge metric, only reflecting the numerator of the *Average Neighbourhood Overlap*, captures the absolute number of shared IP address vertex neighbours between any pair of such vertices. Hence, $G_{Bharti Airtel}$ seems to have more strongly embedded pairs of neighbouring IP address vertices than G_{Aircel} and $G_{Vodafone}$.

Edge metrics by mobile broadband operator at IP granularity			
Mobile broadband operator graph	Average Path Length	Average Neighbourhood Overlap	Average Embeddedness
G_{Aircel}	5.169	.007	.139
$G_{Bharti Airtel}$	4.090	.012	.235
$G_{Vodafone}$	6.404	.006	.138
Key IP: Internet Protocol.			

Table 4-5: Edge metrics by mobile broadband operator at IP granularity.

Having discussed the edge metrics, the next section aims to analyse additional vertex metrics for the three mobile broadband operator networks before examining their Complex Network properties.

4.2.3 Vertex Metric Analysis (IP)

The vertex metrics discussed in this section focus on the properties of IP address vertices in the three mobile broadband operator networks. In detail, we analyse the *Clustering Coefficient* metrics and the *Average Vertex Strength*.

Vertex metrics by mobile broadband operator at IP granularity			
Mobile broadband operator graph	Clustering Coefficient (Directed)	Average Clustering Coefficient	Average Weighted Clustering Coefficient
G_{Aircel}	.009	.009	.009
$G_{Bharti Airtel}$.016	.016	.017
$G_{Vodafone}$.011	.011	.011
Key IP: Internet Protocol.			

Table 4-6: Vertex metrics by mobile broadband operator at IP granularity.

The small *Average Weighted Clustering Coefficients* of the vertices indicate overall weak interaction intensities between the different network IP address vertices.

The plots in Figure 4-2 below illustrate the Average Vertex Strength distributions for the three mobile broadband operator networks. Here, $G_{Vodafone}$ shows considerably stronger weights attaching to edges that links to individual IP address vertices, compared to G_{Aircel} and $G_{Bharti Airtel}$. However, each one of these three networks are showing a uniform distribution of vertex strengths as Figure 4-2 below illustrates.

This evidence shows that the three mobile broadband operator networks display features resembling *Scale-Free Network* models that will be further analysed below (see also section 3.4). Hence, the following section tests the three operator networks against Complex Network properties.

Vertex Strength Distributions at IP granularity

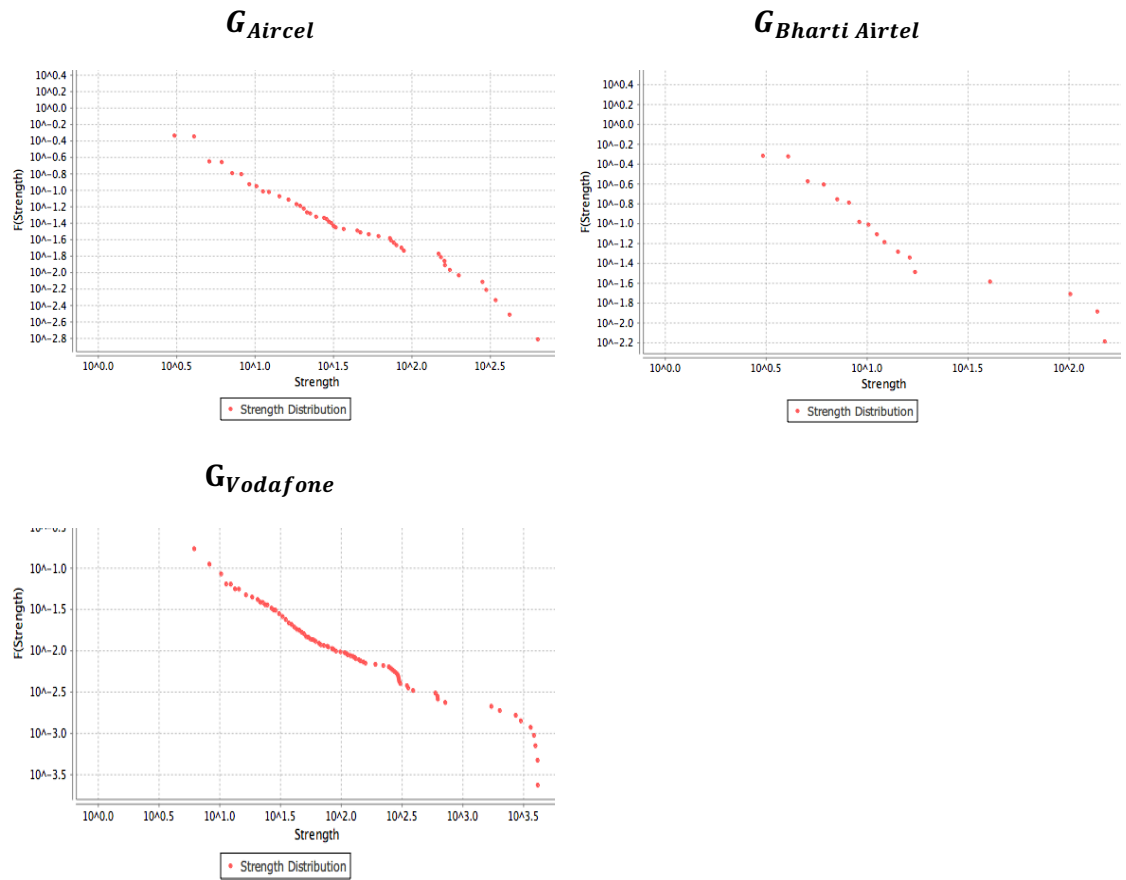


Figure 4-2: Vertex strength distributions by mobile broadband operator at IP granularity.

4.2.4 Complex Network Properties (IP)

The exploration of the networks' edge and vertex metrics shows early evidence of some differences in the internetworking characteristics for the three mobile broadband operator networks. This section tests whether the three operator networks are showing signs of Complex Networks properties as indicated by Boccaletti et al. (2006). Complex Networks refer to non-trivial systems that are usually composed of a vast number of interacting elements with no centralised authority and inherently hard to understand (Strogatz, 2001; Wang, Latapy and Soria, 2012). Such complex systems may be self-organised as their properties emerge from features of vertex interactions (here routing agreements). Self-organisation refers to a complex system where order arises from interactions between previously unordered parts of the system (Bak and Chen, 1991; Wiener, 2014). Complex Networks may come in varying forms such as *Small-World*, *Scale-Free*, *Random*, or *Real-World Network* models. *Small-World Network* models usually show short path lengths as well as high clustering metrics. Similarly, *Scale-Free Network* models also

show short path lengths but would indicate *power-law degree distributions* and little or no clustering (Nefedov, 2013). Agliardi and Giovannetti (1998) study the self-organising criticality and *power-law degree distributions*. Short path lengths may also characterise *Random Networks* but usually incorporate no clustering while following Poisson degree distributions (Vázquez, Pastor-Satorras, and Vespignani, 2002; Nefedov, 2013). Finally, *Real-World Networks* would inhabit short path lengths but also high clustering as well as broad *power-law degree distributions*. Based on Table 4-7 below, we note that the three operator networks indicate relatively small *Average Path Lengths* (below 6.5 hops) and *Clustering Coefficients* (below .017). Figure 4-3 below illustrates the plotted degree distributions indicating linear heavy left-tailed, non-gaussian, *power-law degree distributions*. This suggests evidence for homogeneous network characteristics, rather than heterogeneous ones (Wang, Latapy and Soria, 2012, p.152). We observed, based on the empirical network properties, that the three mobile broadband operator networks display typical signs of *Scale-Free Network* models.

Complex network indicators by operator at IP granularity			
Mobile broadband operator graph	Average Path Length	Clustering Coefficient (Directed)	Degree Distribution
G_{Aircel}	5.169	.009	Power-law
$G_{Bharti Airtel}$	4.090	.016	Power-law
$G_{Vodafone}$	6.404	.011	Power-law
Key IP: Internet Protocol.			

Table 4-7: Complex network indicators by mobile broadband operator at IP granularity.

Power-law Degree Distributions at IP granularity

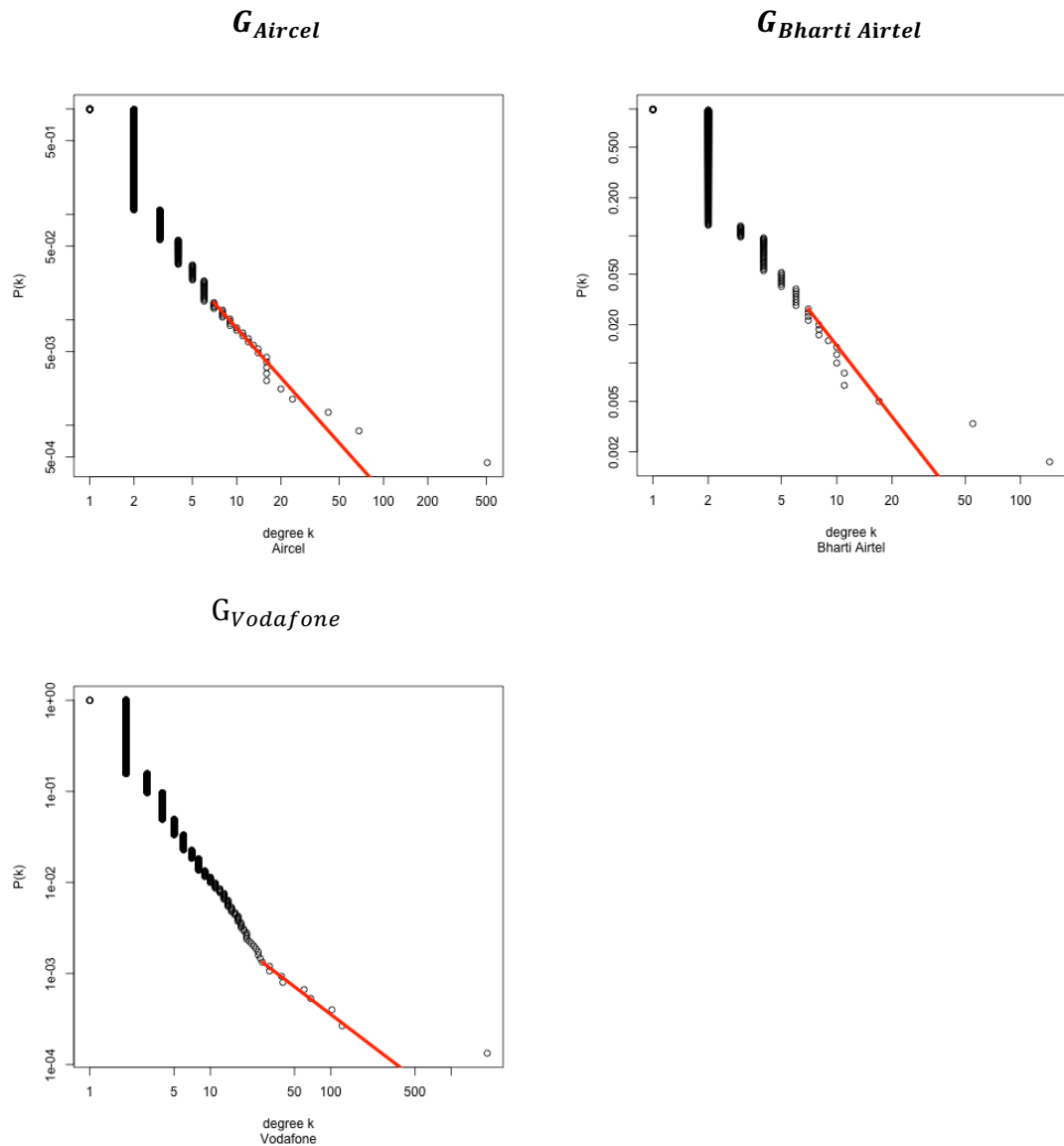


Figure 4-3: Degree distributions by mobile broadband operator at IP granularity.

4.2.5 Summary Network Metric Analysis (IP)

This section analysed the general, edge and vertex metrics of the three Tamil Nadu mobile broadband operator networks. Given the *Average Weighted Degree*, our exploration first implied that Vodafone *traceroutes* traversed recurrent IP addresses far more often than the other two operators, Aircel and Bharti Airtel. Next, the traversed patterns indicated that Vodafone potentially offers a lower end-user's perceived Quality of Service (QoS). Looking next at the *Network Modularity* showed that both Aircel and Vodafone likely have more IP address clusters than Bharti Airtel. Overall, the three mobile operators showed low *Clustering Coefficients* and *Average Weighted Clustering Coefficients*. These metrics revealed weak interaction intensities between different IP addresses in the

upstream Internet market structure. Finally, testing for Complex Network properties showed heavy-tailed *power-law degree distributions*, a characterisation for homogeneous, *Scale-Free Networks*.

Given these findings, the following section aims to test the three mobile broadband operator network's *Scale-Free Network* properties using *Graph Visualisation Analysis*. This should provide us deeper insights into the operators' upstream Internet market structures.

4.2.6 Graph Visualisation Analysis (IP)

For the purpose of exploring the structural operator network properties, we first project the three networks as a graph onto a two-dimensional Euclidean space. Wang, Latapy and Soria, (2012 p.11) note that a first step in *Graph Visualisation Analysis* represents a description of the network structure following by a description of its dynamic evolution. Hence, we look at the resulting network graph visualisations from two different angles. First by using the *Small-World Network* model of Kleinberg (2000), and second by using the *Scale-Free Network* models by Barabási and Albert (2002). Based on these results, we then elaborate a *k-core decomposition* using the algorithm of Alvarez-Hamelin et al. (2005b).

Small-World Network Model (IP)

The *Small-World Network* model by Kleinberg (2000) was used in conjunction with the *Layered Layout* by Kuchar (2012), which is considered to be suitable for *Small-World Network* graph visualisations (see section 2.3.4). To obtain the structural differences of the three operator networks, we consistently chose the same layout properties (see section 3.5.3). The analysis of the edge-distributions in the plotted graph visualisations in Figure 4-4 below indicates that none of the connectivity graphs of the mobile broadband operator networks followed *Small-World Network* properties. However, some IP address vertices are displaying strong relationships to other IP addresses within the circle in the centre of the graph visualisation. Therefore, it seems that the operator networks are showing a core of densely connected IP addresses. This is an indicator for hierarchical upstream Internet market structuring with large *Internet Service Providers (ISPs)* at the core. However, the graph visualisation of Bharti Airtel shows less active connections among a particular set of IP address vertices. This is interesting since it potentially indicates a less hierarchical upstream Internet market structure than the graph visualisations for Aircel and Vodafone.

Moreover, the strongest connected edges are clearly visible on the left-hand side of the graph visualisations (see Figure 4-4 below). From this vertex, the edges seem to leave the visualised *Small-World Network* circle and reach towards other IP address vertices in the network periphery. Moreover, none of the operators' graph visualisations show perfectly interconnected *Small-World Network* effects in the centre of their visualised circles. All graph visualisations seem to build new layers around the centred one, which is most clearly visible for the graph visualisation of $G_{Vodafone}$. This indicates that the networks follow *Scale-Free Network* properties.

Given the findings of the operator's graph visualisations using the *Small-World Network* model with a *Layered Layout*, the next section covers a more detailed analysis and comparison of these networks using the *Scale-Free* Barabási-Albert algorithm in a *Force Atlas 2 Layout*. This algorithm aims to analyse the existence of *Scale-Free Network* properties.

G_{Aircel}

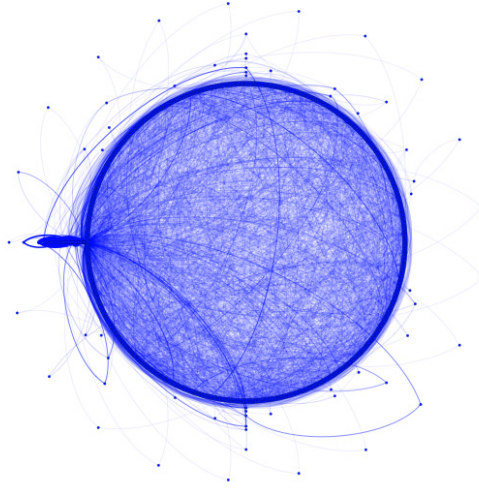
edge-thickness: 0.50.

n – size of lattice: 10.

p – lattice distance to local contacts: 2.

q – long range contacts: 2.

r – clustering exponent: 0.



$G_{Bharti Airtel}$

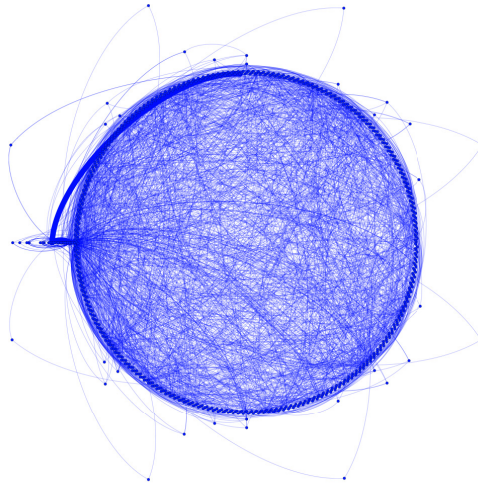
edge-thickness: 0.50.

n – size of lattice: 10.

p – lattice distance to local contacts: 2.

q – long range contacts: 2.

r – clustering exponent: 0.



$G_{Vodafone}$

edge-thickness: 0.25.

n – size of lattice: 10.

p – lattice distance to local contacts: 2.

q – long range contacts: 2.

r – clustering exponent: 0.

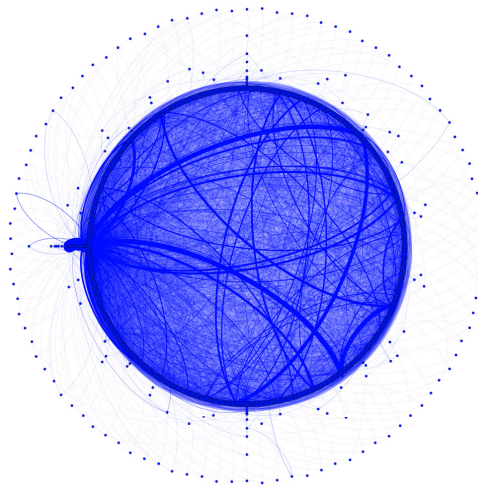


Figure 4-4: Small-World Network graph visualisations in Layered Layout by mobile broadband operator at IP granularity.

Scale-Free Network Model (IP)

Given the findings above, none of the three operator networks seem to display *Small-World Network* properties. Therefore, this section tests the three operator networks by applying *Scale-Free Network* properties using the Barabási and Albert (2002) algorithms. Based on the observed set of IP addresses for each of the three mobile broadband operator networks, we follow these dynamic network procedures to simulate alternative scenarios of network growth emergence. The associated network features will be visualised in comparing the possible evolution of the mobile broadband operator networks below. The utilised algorithms are derived from the work of Barabási and Albert (2002) as stated in Barabási Labs (2013):

- Standard Model with vertex growth and preferential attachment to edges.
- Model A with vertex growth and uniform attachment of edges.
- Model B without vertex growth but preferential attachment to edges.

The mobile broadband operator network graphs were visualised using the *Force Atlas 2 Layout* in the Open Source graph visualisation platform Gephi (2016). This layout is suitable for exploring *Scale-Free Network* properties of networks with up to 10,000 vertices (Jacomy et al., 2014), which none of the three Tamil Nadu mobile broadband operator networks exceeded. The visualisation parameters are stated in section 3.5.3 above. We first simulate and compare the alternative scenarios of network growth emergence of the three mobile broadband operator networks using the Barabási Standard Model (vertex growth and preferential attachment), followed by the Model A with uniform attachment (and retained growth of vertices) and the Model B with preferential attachment ('rich-get-richer' effect) to edges (but no vertex growth).

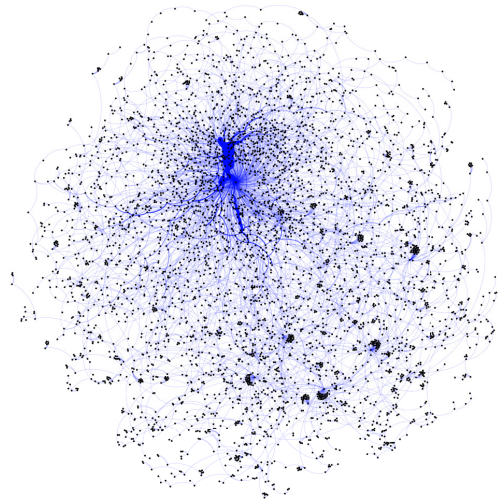
Barabási-Albert Standard Model

First, comparing the three mobile broadband operator networks' graph visualisations at IP granularity using the Barabási-Albert Standard Model indicates structural differences between our three operators of interest. The network simulation considers half of all the networks IP address vertices, being 8,647 simulated vertices in G_{Aircel} , 600 for $G_{Bharti Airtel}$ and 7,509 for $G_{Vodafone}$. To assure comparability, these vertices are chosen based on the total number of vertices in the given operator networks. The Barabási-Albert Standard Model simulation shows that IP address vertices in G_{Aircel} and $G_{Vodafone}$ are more strongly organised in vertex clusters (groupings of IP address vertices) than the vertices in the graph visualisation of $G_{Bharti Airtel}$ (see Figure 4-5 below). Moreover, the

operator networks' graph visualisations of G_{Aircel} and $G_{Vodafone}$ show cores of specific IP addresses that seem densely internetworked. Especially Vodafone seems to make strong use of the same IP address vertices, as indicated by the large blue area in the core of the network. This suggests their potential upstream connectivity reliance on these IP address vertices. On the contrary, the IP address vertices core of $G_{Bharti Airtel}$ seems not strongly internetworked. This is interesting since it indicates an overall fairer distribution of upstream connectivity among the IP address vertices and hence, less internetworking reliance on certain IP address vertices. Moreover, each of the three operator network graph visualisation shows IP address vertices being situated at the edge of the visualised spaces, likely representing IP addresses in the Internet periphery. However, we consider the simulation of network growth emergence using the Barabási-Albert Standard Model to be somewhat fictive, since the IP address vertices, elaborated based on our *Paris traceroute* dataset, represent the unique IP addresses of upstream infrastructure devices (e.g. routers) for the purpose of establishing internetworking connections. Hence, we would not expect a strong vertex growth in a *Real-World Network* growth situation.

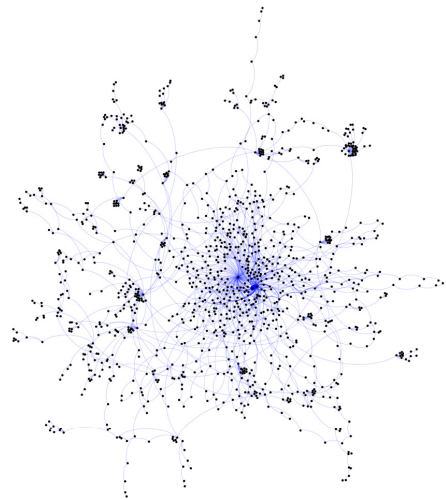
G_{Aircel}

N Number of vertices in network: 8647.



$G_{Bharti Airtel}$

N Number of vertices in network: 600.



$G_{Vodafone}$

N Number of nodes in network: 7509.

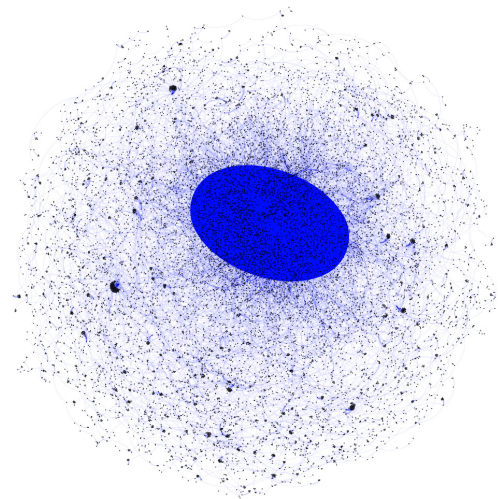


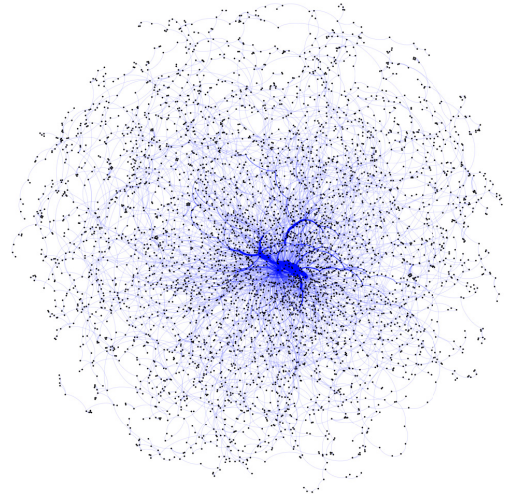
Figure 4-5: Barabási-Albert Standard Model graph visualisations per mobile broadband operator at IP granularity.

Barabási-Albert Model A

Next, we compare the three Tamil Nadu mobile broadband operator networks' graph visualisations at IP granularity using the Barabási-Albert Model A. The network growth simulation of this model considers the growth of vertices in the network but no preferential attachment or 'rich-get-richer' effects. Again, the network simulation of the Barabási-Albert Model A considers 8,647 vertices for G_{Aircel} , 600 for $G_{Bharti Airtel}$ and 7,509 for $G_{Vodafone}$ in the simulation. Interestingly, the three generated network graph visualisations with simulated vertex growth indicate somewhat similar cores of strongly connected IP address vertices. Like at the Barabási-Albert Standard Model above, the graph visualisations of G_{Aircel} and $G_{Vodafone}$ seems to have densely internetworked cores of specific IP addresses. Again, as indicated by the large blue area in the core of the network in Figure 4-6 on the next page, Vodafone seems to make strong use of the same IP address vertices, showing again their potential upstream connectivity reliance on these IP address vertices. On the contrary, the IP address vertices core of $G_{Bharti Airtel}$ seems again not strongly internetworked. Interestingly, compared to the network growth simulation using the Barabási-Albert Standard Model, the simulation of the Barabási-Albert Model A does not indicate the existence of strong IP address vertex clusters. The lack of these vertex clusters may be attributed to the missing preferential attachment of edges. Above, we indicated the somewhat fictive nature of the network growth emergence simulation (given the nature of IP addresses for internetworking and hence upstream Internet connectivity purposes). Adding to this, we indicate the value of the network growth emergence simulation using the preferential attachment of edges, representing connectivity recurrence in a somewhat fix set of upstream IP address vertices. Hence, we consider the following simulation using the Barabási-Albert Model B to be most suitable to understand and graphically analyse the network growth emergence in mobile broadband operator networks.

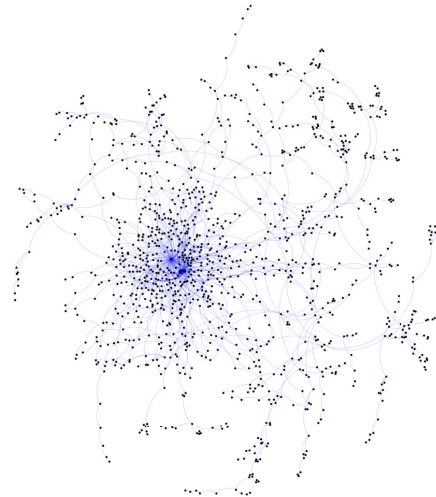
G_{Aircel}

N Number of vertices in network: 8647.



$G_{Bharti Airtel}$

N Number of vertices in network: 600.



$G_{Vodafone}$

N Number of nodes in network: 7509.

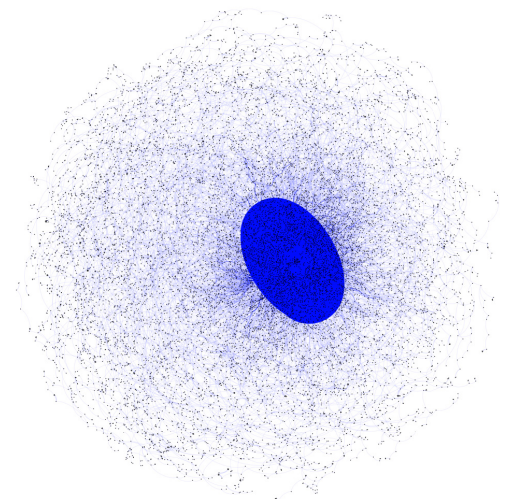


Figure 4-6: Barabási-Albert Model A graph visualisations per mobile broadband operator at IP granularity.

Barabási-Albert Model B

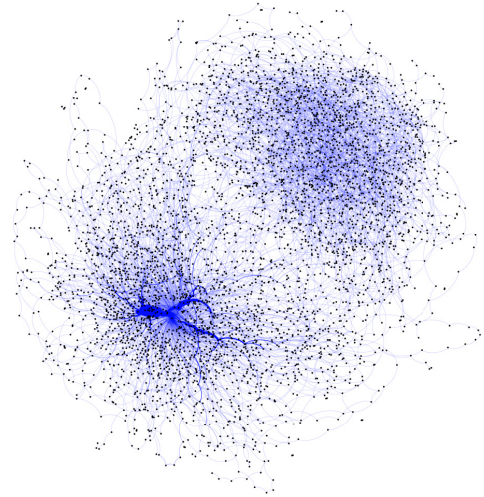
Given the findings above, we indicated that the Barabási-Albert Model B is the most valuable simulation to analyse and understand network growth emergence and structural network properties. Hence, this simulation best represents the connectivity nature of the upstream Internet market, containing a fix set of internetworking-providing agents (Autonomous Systems managing the IP address ranges) but re-establishments of connections amongst the different upstream IP addresses.

Here, the simulated Barabási-Albert Model B again considers the same number of vertices and edges as above, under preferential attachment. The three-generated operator network graph visualisations with simulated vertex growth again indicate somewhat similar cores of strongly connected IP address vertices. In detail, the graph visualisations of G_{Aircel} and $G_{Vodafone}$ have cores of specific IP addresses that are densely internetworked, which is again indicated by the large blue area representing edges between IP address vertices in the following Figure 4-7. Compared to those of G_{Aircel} and $G_{Vodafone}$, the IP address vertices core of $G_{Bharti Airtel}$ are again not strongly internetworked. Interestingly, each of the three operator network graph visualisations using the Barabási-Albert Model B reveals a bi-partitioning of the graph visualisations. This indicates the importance of some IP address vertices that ‘bridge’ connections between the bi-partite parts of the operator networks for internetworking purposes towards the Internet periphery. Here, especially the graph visualisation of $G_{Bharti Airtel}$ indicates the existence of very few of these important vertices. The other two Tamil Nadu mobile broadband operator networks, G_{Aircel} and $G_{Vodafone}$ bridge the apparent bi-partitioning with a multitude of IP addresses, preventing connectivity issues. Moreover, this structuring also indicates a structuring where a few IP address vertices (potentially belonging to larger *Internet Service Providers*) would receive most of the upstream internetworking connectivity, representing connectivity-crucial structural bottlenecks.

G_{Aircel}

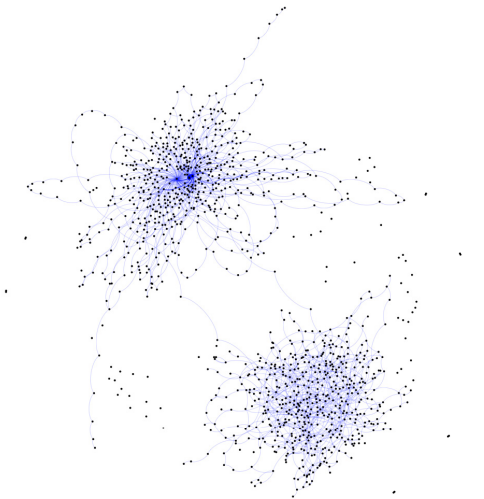
N Number of vertices in network: 8647.

M Number of edges in network: 11411.

 **$G_{Bharti Airtel}$**

N Number of vertices in network: 600.

M Number of edges in network: 803.

 **$G_{Vodafone}$**

N Number of nodes in network: 7509.

M Number of edges in network: 10390.

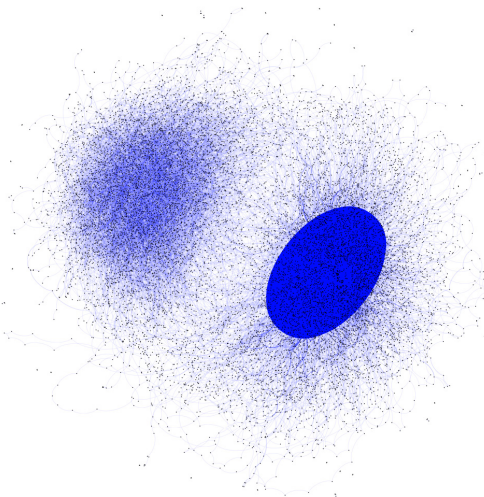


Figure 4-7: Barabási-Albert Model B graph visualisations per mobile broadband operator at IP granularity.

Summarising, this section showed that the Barabási-Albert Model B represents the most valuable simulation to study and understand network growth emergence for traceroute-based mobile broadband operator networks, given the nature of the upstream Internet infrastructure. The three respective graph simulations using the Barabási-Albert Model B then revealed a bi-partitioning of the network graphs. This exposed the structural bottlenecks of certain IP address vertices with an internetworking importance for the three mobile broadband operators, forming a densely-connected core of the operator networks.

Given the findings above, the following section aims to reveal the nature and identity of these influential IP addresses, using the *k-core decomposition* used by Alvarez-Hamelin et al. (2005b) and Busch, Béiro and Alvarez-Hamelin (2011). This will demonstrate the mobile broadband operator's hierarchical upstream Internet market structure.

k-core decomposition (IP)

In this section, we will use the *k-core decomposition* spectral analysis to identify the set of the most densely connected IP address vertices for each of the graphs generated for the three Tamil Nadu mobile broadband operator' networks. Referring to the work of Alvarez-Hamelin et al. (2005b), this *k-core decomposition* reveals the specific roles and relevance of the vertices located in the periphery and core of a network. This method is frequently used for the analysis of Internet structures, such as work of CAIDA shows. Using a *k-core decomposition* algorithm, as introduced by Seidmann (1983), allows for the division of graph visualisation into densely connected network subsets, called *k-cores*. Hence, these *k-cores* represent connectedness properties for the IP address vertices in a given network, where a higher *k-core* indicates a set of more densely connected IP address vertices (see section 2.3.4). The most densely-connected IP address vertices in the network core provide both internetworking connectivity features amongst themselves and between this central core and those IP addresses located in the overall network periphery. Given the identified *k-cores*, this method allows for a clear identification and visualisation of some key hierarchical network properties. Below we start with the *k-core decomposition* for Aircel, followed by the *k-core decomposition* for Bharti Airtel and lastly for Vodafone.

Aircel

When looking at the *k-core decomposition* for the Aircel graph G_{Aircel} in Figure 4-8 below, we observed 179 k-cores. The highest *k-core* is inhabited by three IP address vertices located in the centre of the graph visualisation, indicating the densest connections

amongst these IP address vertices. Using the Maxmind (2015) Geo IP2 database and associating these IP addresses with their Autonomous System Number in Table 4-8 below, we show that all of these three central core IP addresses are associated with Tata Communications (America) Inc. (AS6453), revealing that this Autonomous System plays a key role in providing internetworking connectivity to Aircel to reach the IP address vertices located in the network periphery of the graph generated by the Aircel observations. This supports the previous results that emerged above from exploring the Network metrics. Moreover, we can also identify some vertices that while inhabiting a lower hierarchical *k-core* position, are still providing key connectivity to the periphery.

G_{Aircel}

Highest core IP address vertices, visualised as red vertices in the centre (grey edges):

179 cores: '180.87.39.25'

179 cores: '80.231.154.17'

179 cores: '80.231.217.17'

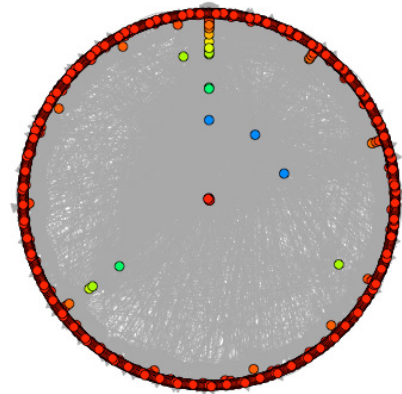


Figure 4-8: Aircel graph visualisation *k-core* decomposition at IP granularity.

Bharti Airtel

Next, when looking at the *k-core decomposition* for the Bharti Airtel graph $G_{Bharti\ Airtel}$ in the following Figure 4-9, we observed 40 *k-cores*. The highest *k-core* is inhabited by two IP address vertices located in the centre of the graph visualisation. This shows the densest connections amongst these IP address vertices, followed by one IP address vertex in the 39th *k-core*. Using the Maxmind (2015) Geo IP2 database and associating these IP addresses with their Autonomous System Number in using the Maxmind (2015) Geo IP2 database and associating these IP addresses with their Autonomous System Number in Table 4-8 below, we show that all of these three central core IP addresses are associated with Bharti Airtel Ltd. (AS45609), Level 3 Communications Inc. (AS3356) and Bharti Airtel Ltd. (AS9498). These Autonomous Systems are playing a key role in providing

Bharti Airtel’s internetworking connectivity to reach the IP address vertices located in the network periphery of the graph, generated by the Bharti Airtel observations. This again supports the previous results that emerged above from exploring the Network metrics. Additionally, we can also identify vertices that, while inhabiting a lower hierarchical *k-core* position, are still providing key connectivity to the periphery.

$G_{Bharti\ Airtel}$

Highest core IP address vertices, visualised as red vertices in the centre (grey edges):

40 cores: ‘223.224.40.92’

40 cores: ‘10.155.84.218’

39 cores: ‘59.144.180.69’

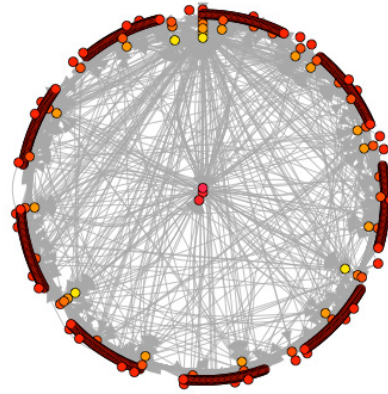


Figure 4-9: Bharti Airtel graph visualisation *k-core* decomposition at IP granularity.

Vodafone

Last, when looking at the *k-core decomposition* for the Vodafone graph $G_{Vodafone}$, as visualised in Figure 4-10 below, we identified 1973 *k-cores*. The highest *k-core* is inhabited by three IP address vertices located in the centre of the graph visualisation, indicating the densest connections amongst these IP address vertices, followed by some IP address vertices in the slightly lower *k-cores*. Using the Maxmind (2015) Geo IP2 database and associating these IP addresses with their Autonomous System Number in using the Maxmind (2015) Geo IP2 database and associating these IP addresses with their Autonomous System Number again in Table 4-8, we show that all of these three central core IP addresses are associated with Vodafone India Ltd. (AS55410). The IP address vertices in the slightly lower central core are associated with the China Education and Research Network Center (AS4538) and Cable and Wireless Worldwide Plc. (AS1273). Similarly, compared to the previous *k-core decompositions*, this again supports the indicated results that emerged above from exploring the Network metrics. Additionally, we can also identify vertices that, while inhabiting a lower hierarchical *k-core* position,

are still providing key connectivity to the periphery.

$G_{Vodafone}$

Highest core IP address vertices, visualised as red, magenta and purple vertices in the centre (grey edges).

1973 cores: '182.19.115.70'

1973 cores: '182.19.114.87'

1973 cores: '182.19.105.88'

1882 cores: '182.19.115.233'

1622 cores: '100.64.0.149'

1458 cores: '166.63.217.41'

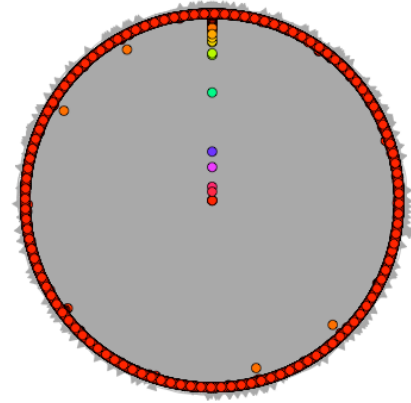


Figure 4-10: Vodafone graph visualisation k -core decomposition at IP granularity.

Highest k-core IP addresses by operator at AS granularity			
Mobile broadband operator graph	k-cores	IP address	Organisational Name (Autonomous System Number) (Maxmind, 2015)
G_{Aircel_AS}	179	180.87.39.25	Tata Communications (America) Inc. (AS6453)
	179	80.231.154.17	Tata Communications (America) Inc. (AS6453)
	179	80.231.217.17	Tata Communications (America) Inc. (AS6453)
$G_{Bharti\ Airtel_AS}$	40	223.224.40.92	Bharti Airtel Ltd. (AS45609)
	40	10.155.84.218	Level 3 Communications Inc. (AS3356)
	39	59.144.180.69	Bharti Airtel Ltd. (AS9498)
$G_{Vodafone_AS}$	1,973	182.19.115.70	Vodafone India Ltd. (AS55410)
	1,973	182.19.114.87	Vodafone India Ltd. (AS55410)
	1,973	182.19.105.88	Vodafone India Ltd. (AS55410)
	1,882	182.19.115.233	Vodafone India Ltd. (AS55410)
	1,622	100.64.0.149	China Education and Research Network Center (AS4538)
	1,458	166.63.217.41	Cable and Wireless Worldwide Plc. (AS1273)
<p>Key</p> <p>AS: Autonomous System.</p> <p>IP: Internet Protocol.</p>			

Table 4-8: Highest k-core vertices by mobile broadband operator at IP granularity.

Summary Graph Visualisation Analysis (IP)

The visualisation analysis allowed us to show that none of the three mobile broadband operator networks displayed *Small-World Network* properties. The *Scale-Free Network* graph simulations provided initial structural insights about the dynamics of network growth emergence based on the existing connectivity. We note that the Barabási-Albert Model B is the most suitable graph algorithm for simulating the emergent network dynamics driven by the possible new connectivity between IP address vertices in the established network based on the principles of the ‘rich-get-richer’ effect. Next, the evidence from the *k-core decomposition* uncovered the most densely connected IP addresses for each of our three Tamil Nadu mobile broadband operator networks. These IP addresses represent structural bottlenecks. Moreover, the mapping of these network core IP address vertices to their associated Autonomous System Numbers revealed structural differences and reliance on different International *Internet Service Providers* between the three mobile broadband operators.

4.2.7 Summary Complex Network Analysis (IP)

The first part of the *Complex Network Analysis* included the *Descriptive Analysis* of the complete operator networks, covering all *traceroute* hop observations at Internet Protocol granularity. We indicated the relevant observations in this case study and analysed their general, edge and vertex metrics. Moreover, we found that Vodafone had IP addresses that are considerably more often traversed than those of Aircel and Bharti Airtel. Based on the *Clustering Coefficient* analysis, all three mobile broadband operator networks indicated low vertex interaction intensities but also *power-law degree distributions* and likely *Scale-Free network* features. This evidence was further tested in the graph visualisation simulations, where the BA-Model B showed the existence of densely connecting cores for each operator network. The apparent cores of the *Scale-Free Network* graph visualisations were then explored by using the *k-core decomposition*, which revealed the most important IP address vertices that each operator network relied upon for internetworking connectivity to the network periphery. More precisely, Aircel showed a reliance on Tata Communications (America) Inc. Bharti Airtel showed a strong reliance on its own Autonomous Systems as well as Level 3 Communications Inc. Lastly, Vodafone also strongly relied on of their own Autonomous Systems, including their Cable and Wireless Worldwide plc. subsidiary, but also on the Chinese Research and Education Network. Although these findings are very interesting, we assume that the

Complex Network Analysis at IP granularity hides crucial upstream internetworking connectivity between *Internet Service Providers*. One could assume that the same analysis at AS granularity illustrates additional upstream Internet market structural properties, where the identified IP addresses belong to the same set of Autonomous Systems. Hence, the following section covers our analysis at AS granularity. This granularity is not only closer to an organisational level but also represents the level on which *Internet Service Providers* arrange economic connectivity relationships between Autonomous Systems.

4.3 Complex Network Analysis at AS granularity

4.3.1 Data Preparations

Here we describe the necessary preparations to further explore the previously identified structural features of the three operator networks from a higher level of perspective, the Autonomous System granularity. Hence, the key step performed in this section was to transform every IP address vertex observation into their Autonomous System Number. Autonomous Systems (ASs) are managed by either one, or a cooperating multitude of, *Internet Service Providers*. Moreover, each Autonomous System obtains a unique identifier, its Autonomous System Number (ASN), which are allocated and administered by the Internet Assigned Numbers Authority (IANA, 2016). Here an *ISP* registers the respective AS Number for the purpose of Border Gateway Protocol (BGP) routing. IANA then assigns the AS Number to the responsible Regional Internet Registry (RIR), which subsequently assigns the registered ASN to the applying *Internet Service Provider*. Once obtained, the AS Number then represents a collection of one, or a multitude of, IP address prefixes following the Class Inter-Domain Routing (CIDR) notation. This means that an AS Number manages a range of IP addresses. Hence, an *ISP* not only obtains the registered AS Number, but also the (single) administrative control over the Autonomous System and its associated IP addresses. This process generates the entire Internet address space that is required to establish Inter-domain routing policies. An Autonomous System may find these routing policies in an operators' routing tables being stored on its own routers and providing the key routing data for BGP connectivity instructions. Hence, different *Internet Service Providers* usually publish and share these routing tables for (bilateral) connection purposes.

To obtain the Autonomous System Numbers of the 33,388 collected unique IP addresses-based *traceroute* hop observations, we fused these IP addresses with the secondary Maxmind (2015) Geo IP2 dataset. This transformation from IP address to ASN represents a crucial part of the upcoming *Complex and Statistical Network Analysis* at AS granularity. In this way, we were able to associate to every IP address the corresponding Autonomous System Number (e.g. ‘AS174’) and name and headquarter location of the Internet Service Provider managing the ASN (e.g. ‘Cogent Communications’, ‘USA’). The obtained AS granularity is of particular importance when looking at the economic relationships between the upstream Autonomous Systems in the three mobile broadband operator networks. Our mapping results were, given their importance for the analysis in the upcoming chapters, verified by using other credible sources, including UltraTools (2015), Hurricane Electric (2016) and Team Cymru (2016). The overall correct mapping justified the suitability of the secondary Maxmind (2015) dataset. After transforming the IP address set of relationships of each mobile broadband operator networks into a set of Autonomous System Number relationships, we imported the set of these AS relationships as directed edge-table into Gephi (2015). This procedure allowed us to generate the three Tamil Nadu mobile broadband operator networks at AS granularity. The following Table 4-9 below compares the number of vertices and edges at both granularities. This indicates the effects of the mapping described above. Unsurprisingly, the mapping resulted in a much smaller number of unique Autonomous System vertices and edges linking those AS vertices. This reflects the nature of Autonomous Systems operating IP address prefix range(s).

Comparison between vertices and edges at IP and AS granularity per mobile broadband operator				
Mobile broadband operator	Vertices (IP)	Vertices (AS)	Edges (IP)	Edges (AS)
Aircel	2,259	522	2,879	1,144
Bharti Airtel	600	180	803	388
Vodafone	7,509	1,513	10,390	3,627
Key AS: Autonomous System. IP: Internet Protocol.				

Table 4-9: Vertices and edges at IP and AS granularity per mobile broadband operator.

The graph for Aircel was now denoted as G_{Aircel_AS} (522,1144), the one of Bharti Airtel as $G_{Bharti\ Airtel_AS}$ (180,388) and the Vodafone one as $G_{Vodafone_AS}$ (1513,3627). Here, we added a ‘_AS’ suffix, which helps to differentiate between the two granularities of analysis.

4.3.2 General Network Metric Analysis (AS)

Just like at IP granularity, we analysed the *Average (Weighted) Degree*, the *Network Diameter*, the *Network Density* and the randomized *Modularity*.

Next, Table 4-10 below indicates that $G_{Vodafone_AS}$ showed a higher *Average Weighted Degree* of 20.247 than the other two operators. Following, we report the calculated values for the three mobile broadband operator graphs’ *Network Diameter* (directed), a metric that indicates the longest possible shortest distance paths between any two Autonomous System vertices in the graph (see section 3.4.2). Table 4-10 below lists that the graph of the Vodafone observations, $G_{Vodafone_AS}$ showed, again, the greatest *Network Diameter* (12). Given these values, we add to our previous statement at IP granularity (see section 4.2.1 above) that the larger *Network Diameter* of $G_{Vodafone_AS}$ shows longer possible shortest internetworking distance paths in the upstream Internet connectivity.

General metrics by mobile broadband operator at AS granularity					
Mobile broadband operator graph	Average Degree	Average Weighted Degree	Diameter (Directed)	Density (Directed)	Modularity (randomized)
G_{Aircel_AS}	2.192	9.098	6	.004	.271
$G_{Bharti\ Airtel_AS}$	2.156	5.331	8	.012	.337
$G_{Vodafone_AS}$	2.397	20.247	12	.002	.293
Key AS: Autonomous System.					

Table 4-10: General metrics by mobile broadband operator at AS granularity.

The overall small values of the mobile broadband operators’ *Graph Density* (directed) might indicate that Autonomous System vertices are not densely connecting between each other, giving a potential indicator for *power-law degree distributions* and a *Scale-Free Network* model as shown in section 4.2.4 above. Hence, we expect that a small number

of Autonomous Systems would show many connections to other Autonomous Systems in the network. Table 4-10 above states the *Graph Densities*. We assume that the overall increase in the *Graph Densities*, compared to the ones at IP granularity, results from mapping the IP addresses to Autonomous Systems.

Lastly, the *Randomised Modularity*, a relevant metric to capture structure and dynamics of a network (Newman, 2006), revealed sparser connections compared to those at IP granularity. Our data showed the existence of 117 vertex clusters in $G_{Vodafone_AS}$, 96 in G_{Aircel_AS} and only 35 in $G_{Bharti\ Airtel_AS}$. Hence, we assume that, at Autonomous System granularity, $G_{Vodafone_AS}$ and G_{Aircel_AS} are more likely to organise in dense connections between vertices in clusters but sparse connections between vertices linking those clusters. The transformation of our three Tamil Nadu mobile broadband operator networks from IP to AS granularity, therefore, resulted in a loss of community structure properties. This shows that IP addresses that previously tended to be organised in IP address clusters are now subsumed in Autonomous Systems. Moreover, these values indicate the potential reliance on different Autonomous Systems involved in internetworking amongst other Autonomous Systems in the operator networks.

In the next section, we will analyse additional edge metrics with the objective of uncovering more structural features for our three mobile broadband operator networks at Autonomous System granularity, while comparing the evidence to those at Internet Protocol granularity.

4.3.3 Edge Metric Analysis (AS)

The edge metrics discussed in this section focus on the connections between any given set of Autonomous System vertices in our three mobile broadband operator networks. Here, we analyse the *Average Path Length*, the *Edge Betweenness*, the *Average Neighbourhood Overlap* and the *Average Embeddedness* edge metrics, as reported in Table 4-11 below.

Edge metrics by mobile broadband operator at AS granularity				
Mobile broadband operator graph	Average Path Length	Edge Betweenness	Average Neighbourhood Overlap	Average Embeddedness
G_{Aircel_AS}	2.423	8,782	.227	1.353
$G_{Bharti\ Airtel_AS}$	2.192	2,142	.232	1.387
$G_{Vodafone_AS}$	4.563	305,742	.208	1.772
Key AS: Autonomous System.				

Table 4-11: Edge metrics by mobile broadband operator at AS granularity.

Compared to the edge metrics at IP granularity, the *Average Path Length* shortened for all three mobile broadband operator networks at AS granularity, representing a data fusion effect (see Table 4-11 above and section 3.4.2). These differences still indicate that Bharti Airtel routes connections more efficiently, or that Vodafone and Aircel more heavily rely on Autonomous Systems' internal routings.

The second analysed edge metric at Autonomous System granularity was the *Average Neighbourhood Overlap* (see Table 4-11 above), where the highest value was again shown by $G_{Bharti\ Airtel_AS}$. This suggests that the average connections of any pair of neighbouring Autonomous System vertices in $G_{Bharti\ Airtel_AS}$ are not well connected among themselves. On the contrary, the average connections of any pair of neighbouring Autonomous System vertices in $G_{Vodafone_AS}$ are well connected among themselves. Furthermore, these findings are, again, supported by those of the *Average Embeddedness*, where the highest *Average Embeddedness* value was shown by $G_{Vodafone_AS}$. Therefore, $G_{Vodafone_AS}$ seems to have more strongly embedded pairs of neighbouring Autonomous System vertices than the other two mobile broadband operator networks.

The *Edge Betweenness*, which looks at the number of shortest paths going through certain Autonomous System vertices, through an edge, in a given network provides some interesting structural insights. Here, the highest *Edge Betweenness* value was shown by $G_{Vodafone_AS}$. Hence, we assume the existence of some very strong and constantly traversed edges in $G_{Vodafone_AS}$, likely to be Vodafone's own Autonomous Systems. Moreover, this points towards the existence of structural bottlenecks in the upstream

Internet market for $G_{Vodafone_AS}$ and G_{Aircel_AS} .

Having discussed the edge metrics of the three mobile broadband operator networks at Autonomous System granularity, the next section aims to analyse additional vertex metrics for the operator networks before examining their Complex Network properties.

4.3.4 Vertex Metrics (AS)

Following the analysis of edge metrics above, we analyse the *Clustering Coefficient*, the *Average Clustering Coefficient*, the *Average Weighted Clustering Coefficient* and the *Average Vertex Strength*, as indicated in the following Table 4-12.

Again, these (*Average*) *Clustering Coefficient* are of particular interest since they indicate whether the neighbours of Autonomous System vertices in the given mobile broadband operator networks are actively connecting between themselves. The (*Average*) *Clustering Coefficient* revealed some differences between the three operator networks. These three *Clustering Coefficient* values in Table 4-12 below indicate the existence of mutual internetworking connectivity between Autonomous Systems and their given neighbouring vertices.

High *Clustering Coefficient* values indicate dense connections, likely to show *Small-World Network* properties. Given the somewhat low *Clustering Coefficients* of the three mobile broadband operator networks, we again assume the presence of *Scale-Free Network* models. This feature is analysed by looking at the Complex Network properties and the *Graph Visualisation Analysis* in the following sections.

Vertex metrics by mobile broadband operator at AS granularity			
Mobile broadband operator graph	Clustering Coefficient (Directed)	Average Clustering Coefficient	Average Weighted Clustering Coefficient
G_{Aircel_AS}	.105	.167	.099
$G_{Bharti Airtel_AS}$.085	.155	.081
$G_{Vodafone_AS}$.170	.257	.165
Key AS: Autonomous System.			

Table 4-12: Vertex metrics by mobile broadband operator at AS granularity.

The three plots in the following Figure 4-11 illustrate the *Average Vertex Strength* distributions for the three mobile broadband operator networks at AS granularity. This distribution represents the average sum of weights being attached to edges (incoming and outgoing connections) belonging to an Autonomous System vertex (Barrat et al., 2004). The visualised the *Average Vertex Strength* distributions, using R (2016) as indicated in see section 3.4.2, support the previous findings at IP granularity. In detail, $G_{Vodafone_AS}$ and G_{Airtel_AS} are showing stronger vertex strengths of a few Autonomous Systems (bottom-right of the distribution plots) that seem to links to the many with considerably less strong vertex strengths, compared to $G_{Bharti_Airtel_AS}$. In detail, $G_{Vodafone_AS}$ reveals a *Average Vertex Strength* of 40.493 compared to G_{Airtel_AS} with 19.141 and $G_{Bharti_Airtel_AS}$ with an *Average Vertex Strength* of 10.622.

Vertex Strength Distributions at AS granularity

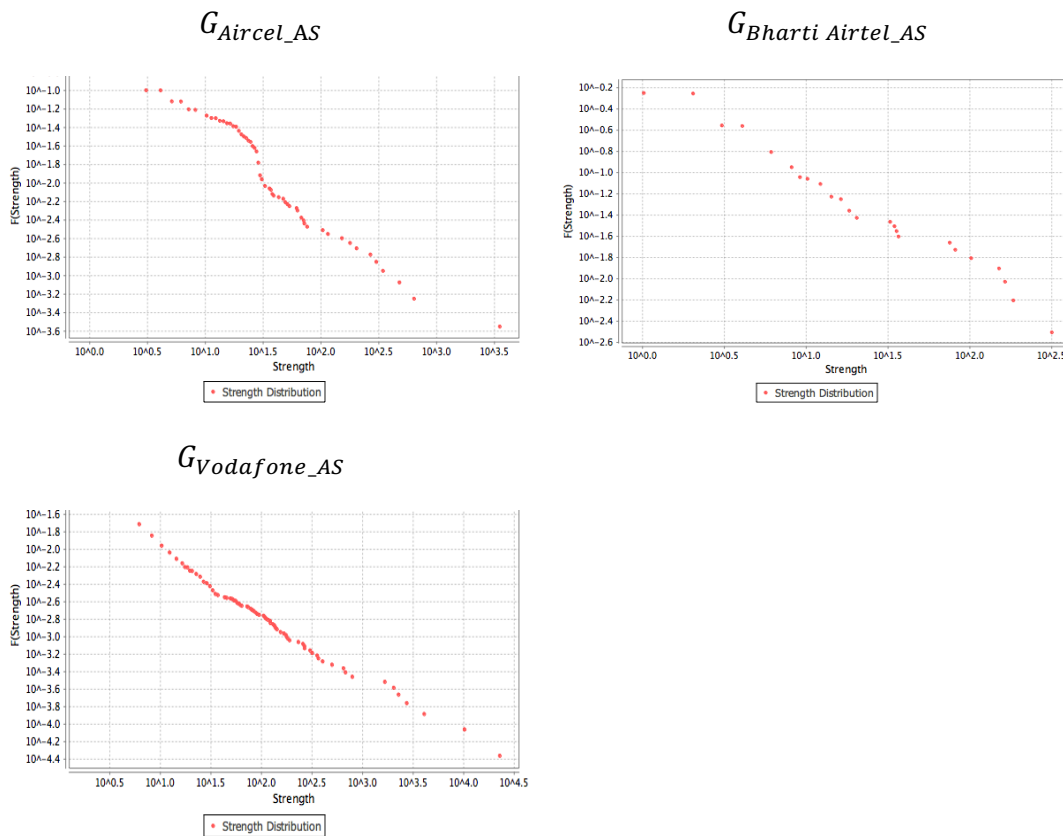


Figure 4-11: Vertex strength distribution by mobile broadband operator at AS granularity.

This section analysed the general, edge and vertex metrics of the three Tamil Nadu mobile broadband operator networks at Autonomous System granularity. Given the *Average Weighted Degree*, our exploration implied that Vodafone *traceroutes* traversed recurrent

Autonomous Systems far more often than the other two mobile broadband operators. Looking next at the *Network Modularity* showed that both Vodafone and Aircel were more likely to organise in Autonomous System clusters than Bharti Airtel. The *Edge Betweenness* showed that Vodafone had edges that were constantly traversed, indicating the structural importance of a few Autonomous Systems. The *Average Neighbourhood Overlap* indicated that Vodafone had AS vertices being well connected among themselves. Overall, the three mobile operators exposed a potentially low (Average) *Clustering Coefficients*, indicating weak interaction intensities between different Autonomous Systems in the upstream Internet market structure and potentially pointing towards a *Scale-Free Network* nature and hierarchical structuring.

4.3.5 Complex Network Properties (AS)

The results of the operator network metrics seen above reveal some properties of structural relevance of some Autonomous System vertices. These Autonomous Systems would, therefore, inhabit positions that are likely to indicate *Scale-Free Network* properties, rather than *Small-World Network* ones. This section summarises a number of metrics and assesses the networks against their Complex Network properties. The next section generates the respective graph visualisations at AS granularity, while comparing their properties to those at IP granularity. Expanding on the previous analysis, Table 4-13 below reports the obtained values.

Complex Network Indicators by mobile broadband operator at AS granularity			
Mobile broadband operator graph	Average Path Length	Clustering Coefficient (Directed)	Degree Distribution
G_{Aircel_AS}	2.423	.105	Power-law
$G_{Bharti\ Airtel_AS}$	2.192	.085	Power-law
$G_{Vodafone_AS}$	4.563	.170	Power-law
Key AS: Autonomous System.			

Table 4-13: Complex network indicators by mobile broadband operator at AS granularity.

While all of the possible network models (*Random*, *Real-World*, *Small-World* and *Scale-Free*) imply, the existence of short *Path Lengths* the first differences emerge when looking at the *Clustering Coefficients*. While *Random Networks* usually show no *Clustering Coefficient*, following the notion of Watts and Strogatz (1998), *Real-* and *Small-World Networks* indicate a somewhat high, or higher, *Clustering Coefficient*, as Table 4-14 indicates. *Scale-Free Network* models on the other hand show either no, or very little clustering (Nefedov, 2013). *Random Networks* represent the only network model that follows a *Poisson distribution*, whereas *Small-World Networks* show overall constant *Degrees*, considering that every vertex in the network should be very well connected to all the other vertices. Moreover, *Real-World Networks* usually indicate heavy tails in their *Degree* distributions. Lastly, *Scale-Free Networks* are showing also heavy tailed, but *power-law degree distributions*.

Comparison of complex network model properties			
Network Model	Path Lengths	Clustering Coefficient	Degree distribution
Random	short	no	Poisson distribution
Real-World	short	high	Heavy tails, often power-laws
Small-World	short	high	(nearly) constant
Scale-Free*	short	little or none	Heavy tails, power-laws
Key * suitable network properties.			

Table 4-14: Comparison of complex network model properties, Source: Nefedov (2013).

Remembering that the *Average Path Length* for each of the graphs considerably dropped after the mapping from IP to Autonomous System granularity, we can still classify the *Path Lengths* as short (below 4.6). Next, we note that the three operator networks are still indicating small *Clustering Coefficients* (below .180). Figure 4-12 below represents the plot of the *Degree* distributions and exposed linear, heavy-tailed, non-Gaussian *power-law degree distributions*. The nature of these *Degrees* could either be strongly incoming connections, strongly outgoing ones, or a combination of the two. Of special interest is the number of vertices with a fairly small *Degree* on the left-hand side of the plots in the

following Figure 4-12. This means that most Autonomous Systems in the operator networks have low connectivity *Degrees* and hence a small number of incoming or outgoing interworking connections. Moreover, those Autonomous System vertices with a higher *Degree* (k) on the right-hand side of the illustrated plots indicate the structural importance of certain Autonomous Systems, or their *Internet Service Providers*, for the connectivity of the three mobile broadband operator networks.

Power-law Degree Distributions at AS granularity

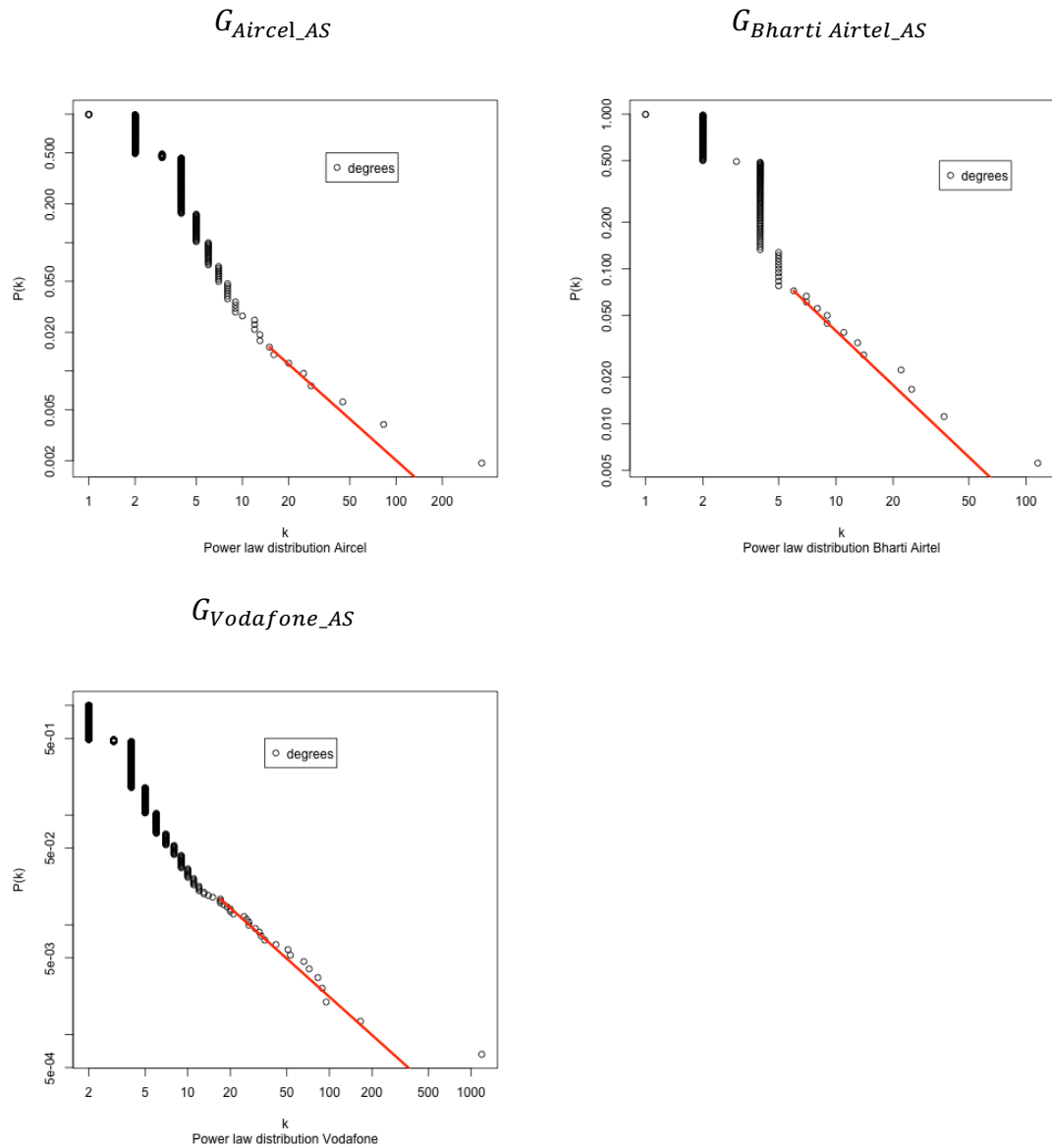


Figure 4-12: Degree distributions by mobile broadband operator at AS granularity.

Based on the obtained indicators, we observed that the three Tamil Nadu mobile broadband operator networks are displaying *Scale-Free Network* properties.

These features are further analysed in drawing the respective graph visualisations using

both *Small-World Network* and *Scale-Free Network* visualisation, and simulation, algorithms in the following section, after summarising the *Complex Network Analysis* at Autonomous System granularity.

4.3.6 Summary Network Metric Analysis (AS)

The general network metrics indicated that the Vodafone network incorporates ASes that are more often traversed than those in the other two operator networks. Analysing the edge metrics led to the discovery that the Vodafone network is either less efficiently organised than the other two mobile broadband operators, or that they are constantly making use of their own Autonomous Systems for upstream internetworking connectivity purposes. While displaying that individual ASes are traversed more often than others, the edge metrics also inferred that the operator networks incorporate systematic organisation of relationships between Autonomous Systems. This points towards a hierarchical upstream Internet market structuring, rather than a random one. Unsurprisingly then, the vertex metrics revealed higher interaction intensities as well as stronger Autonomous System vertex strengths in the respective operator network graphs. The metrics also showed that Vodafone makes use of their own Autonomous Systems, while Bharti Airtel had a strong reliance on other Autonomous Systems, both of which are indicators for potential presence of upstream connectivity bottlenecks. Moreover, by looking at the Complex Network properties, we suggest that the mobile broadband operator networks are following *Scale-Free Network* models. These features are analysed using *Graph Visualisation Analysis* and simulations of the emerging features of network evolution, parameterised on current connectivity evidence, in the next section.

4.3.7 Graph Visualisation analysis (AS)

Small-World Model (AS)

The Network metrics above indicate that the three Tamil Nadu mobile broadband operator networks are following *Scale-Free Network* properties. Therefore, this section aims to challenge this indication by testing the three operator networks, again, against their *Small-World* and *Scale-Free Network* properties.

Here, the three mobile broadband operator networks were, again, placed into a two-dimensional Euclidean Space, following the *Layered Layout* by Kuchar (2012), while the visualisation parameters are stated in the sections 3.5.3 and 3.5.5 above. Figure 4-13 below shows that none of the three mobile broadband operator graph visualisations

signals *Small-World Network* properties. However, some Autonomous System are showing a core of densely connected Autonomous Systems, an indicator for hierarchical upstream Internet market structuring with large *Internet Service Providers*. Again, the graph visualisation of Bharti Airtel shows less active connections among a particular set of Autonomous System vertices (similarly to IP granularity). This is interesting since it indicates a less hierarchical upstream Internet market structure than the graph visualisations for Aircel and Vodafone. Moreover, the strongest connected edges are, again, clearly visible on the left-hand side of the graph visualisations, which becomes most apparent for the network graph visualisation of Vodafone. From the highly connected vertices displayed in the following Figure 4-13, the edges leave the *Small-World Network* circle and reach towards other vertices in the upstream Internet periphery. Additionally, none of the operators' graph visualisations shows perfectly interconnected *Small-World Network* effects in their visualised graph circles. All graph visualisations seem, hence, to build new layers around the centred one, which is again most clearly visible for the graph visualisation of $G_{Vodafone_AS}$. This indicates, as anticipated, that the mobile broadband operator networks follow *Scale-Free Network* properties.

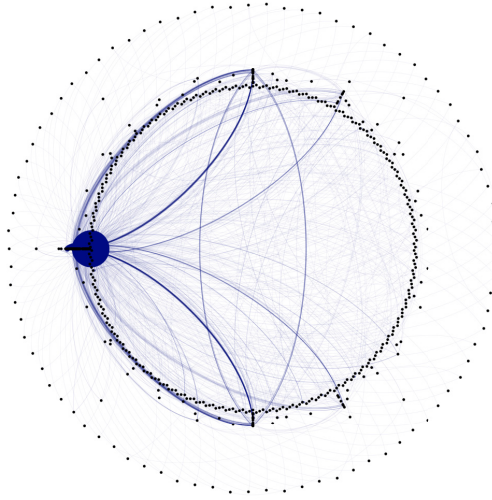
G_{Aircel_AS}

n – size of lattice: 10.

p – lattice distance to local contacts: 2.

q – long range contacts: 2.

r – clustering exponent: 0.

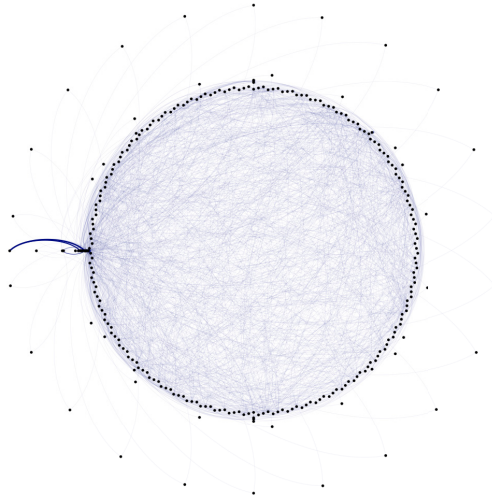
 **$G_{Bharti_Airtel_AS}$**

n – size of lattice: 10.

p – lattice distance to local contacts: 2.

q – long range contacts: 2.

r – clustering exponent: 0.

 **$G_{Vodafone_AS}$**

edge-thickness: 0.2

n – size of lattice: 10

p – lattice distance to local contacts: 2

q – long range contacts: 2

r – clustering exponent: 0

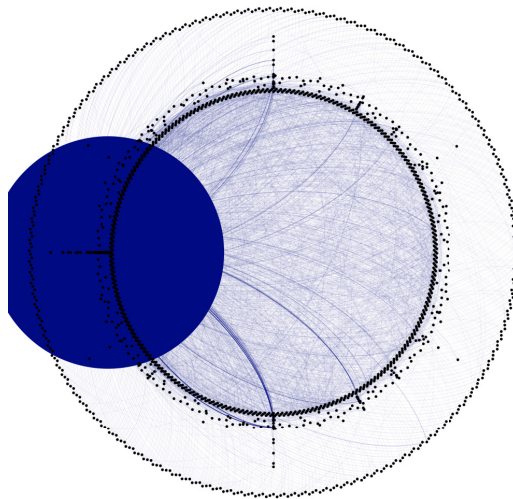


Figure 4-13: Small-World Network graph visualisations in Layered Layout by mobile broadband operator at AS granularity.

Scale-Free Barabási-Albert Model B (AS)

Next, we assess the fit of the three mobile broadband operator networks to the *Scale-Free Network*. Following the indications as IP granularity we visualise, our three mobile broadband operator network graphs only using the Barabási-Albert Model B, since the other two simulations were less useful for simulating and interpreting network growth emergence for upstream Internet connectivity between IP addresses.

Our analysis at Autonomous System granularity captures again a visual analysis of the generated network graph simulations and compares the differences first amongst the mobile broadband operators and second, to the simulated graph visualisations at Internet Protocol granularity. We used again the same visualisation and simulation properties for the network growth emergence (see section 3.5.3). For readability purposes, we set the edge-thicknesses independently, where the graph visualisation of $G_{Vodafone_AS}$ needed a very low edge-thickness to avoid edges occupying the complete visualisation space.

Our dynamic network procedures simulate the graph visualisation for network growth using the preferential attachment of edges, meaning that the more connected Autonomous System vertices were more likely to receive new edges in the growth model. Additionally, vertices with a higher *Degree* would have a stronger ability to grab preferentially attached links, creating a scenario of mobile broadband operator network evolution. This preferential attachment seemed to suit the nature of *traceroutes* since it represents new connectivity patterns that emerge given the more or less fix number of 49,874 Autonomous System vertices during the data collection campaign, see CAIDA (2016a). The latest CAIDA (2016a) AS-Rank dataset (May 2016) shows 54,722 Autonomous Systems, meaning that some vertex growth in the network simulation would be justified, but we did not have the ability to change vertex growth settings in the employed Gephi (2016) simulation algorithm. Nevertheless, the generated graph visualisations using the Barabási-Albert Model B represent a valuable approach to expose important parts of the network structure, given its edge evolution under preferential attachment.

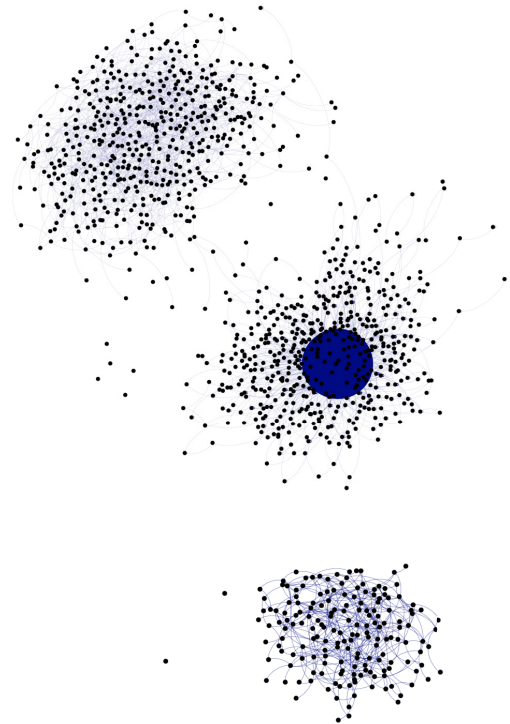
Here, the Barabási-Albert Model B simulation considers 522 Autonomous System vertices (without growth) for G_{Aircel_AS} , 180 for $G_{Bharti_Airtel_AS}$ and 1,513 for $G_{Vodafone_AS}$, covering the full set of AS vertices as Table 4-9 before lists. The three-generated operator network graph visualisations with simulated Autonomous System vertex growth indicate densely connected AS vertex cores for each of the three mobile

broadband operators, as shown by the large blue area representing edges between vertices in Figure 4-7 on the next page. This structuring indicates that a few Autonomous System vertices (potentially belonging to larger *ISPs*) likely receive most of the upstream internetworking connectivity, representing connectivity-crucial structural bottlenecks. This was already indicated at IP granularity. Especially $G_{Vodafone_AS}$ shows only very few Autonomous Systems that are highly internetworking with each other, indicated by the ‘fat’ blue edges in Figure 4-7 (using layout parameters, see sections 3.5.3 and 3.5.5). Moreover, each of the three operator network graph visualisations reveals a stronger bi-partitioning of the visualised operator network graphs, compared to those at IP granularity. The number of connectivity- ‘bridging’ vertices between the bi-partite parts of the operator networks, potentially connecting the Internet periphery, became even more apparent than at IP granularity. Here, G_{Aircel_AS} indicates a very small number of these bridging AS vertices, whereas $G_{Bharti_Airtel_AS}$ loses them, in this simulation, almost entirely. Also $G_{Vodafone_AS}$, considering the larger number of observations, only shows very few of these structurally important Autonomous Systems. Overall, $G_{Vodafone_AS}$ seems to have more Autonomous Systems in the Internet periphery that are less well connected to the core than those in G_{Aircel_AS} or $G_{Bharti_Airtel_AS}$, indicated by the thinner edges between the AS vertices.

G_{Aircel_AS}

N Number of nodes in network: 522.

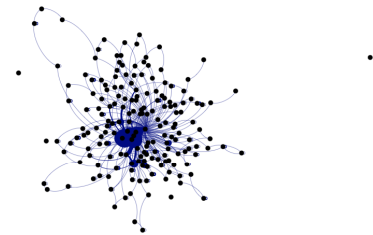
Edge-thickness: 0.25.



$G_{Bharti_Airtel_AS}$

N Number of nodes in network: 180.

Edge-thickness: 1.



$G_{Vodafone_AS}$

N Number of nodes in network: 1,513.

Edge-thickness: 0.05.

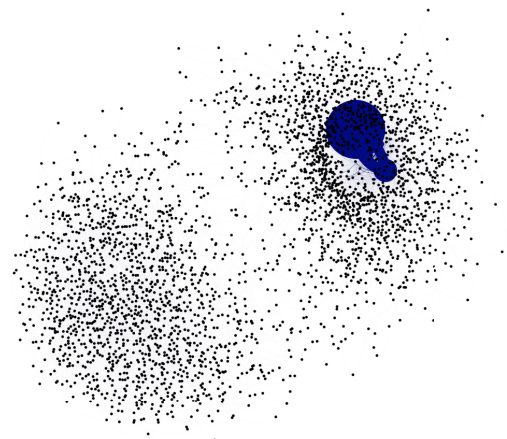


Figure 4-14: Barabási-Albert Model B graph visualisations per mobile broadband operator at AS granularity.

The BA-Model Model B visualisations in Figure 4-14 perfectly indicate the effect of the transformation from IP to AS granularity. While the strong cores are still visible in the centre of the graph visualisations, the vertex hubs and structural gaps become even more apparent. Interestingly, the single cores in the graph visualisations of $G_{Vodafone_AS}$ and G_{Airtel_AS} indicate a particularly strong connection to very few Autonomous Systems. On the other hand, $G_{Bharti_Airtel_AS}$ shows a number of somewhat stronger connections among a multitude of Autonomous Systems. The preferential attachment algorithm of the BA-Model increases the structural indicators but might generally blur the importance of single vertices, or their respective relationships. Nevertheless, we consider these graph visualisations as a very important analytical step towards the identification of market structural properties, as it provides some indicators for certain structures in the respective operator networks.

To further uncover the properties of upstream Internet market structures, next we will look at the operator network's centrality measurements. These measurements should reveal Autonomous Systems with particularly interesting network structuring properties.

Centrality Metrics (AS)

In this section, we focus on exploring the centrality measurements for each of the three operator networks, to gain a better understanding of their structural network properties.

First, the *Degree Centrality* estimates the importance of a vertex based on its obtained *Degrees* (*In-* and *Out-Degree*), following a linear interpretation of the relevant connections. Second, the *Closeness Centrality* represents the sum of the lengths of the shortest paths between a certain Autonomous System vertex and all other vertices in a network. More central Autonomous Systems are therefore closer (not in geo-distance) to all other operator network vertices. Thirdly, the *Betweenness Centrality* explains the centrality of an Autonomous System vertex based on the number of shortest paths passing through a certain vertex. Lastly, the *Eigenvector Centrality* measures vertex influence within the operator network, calculated based on the concept that edges to high-scoring vertices contribute more to a given vertex influence than edges to low-scoring ones. The most central vertices in a network are therefore those of highest influence, representing the most valuable Autonomous Systems to connect to.

Moreover, we extend each Autonomous System with its associated Transit Degree from the secondary CAIDA (2016a) AS-Rank dataset. This Transit Degree measurement

represents the number of Autonomous Systems for which a certain Autonomous System was observed to receive transit paths in the list of all Autonomous Systems in the CAIDA (2016a) AS-Rank dataset. An Autonomous System received transit path also indicates those downstream ones that pay transit fees to the receiving AS. Transit itself means that an Autonomous System provides connectivity to the publicly available Internet routing tables, whereas payments are made upon traffic volume-based fees. This is considerably different to peering, where traffic is bilateral between a pair of peering Autonomous Systems (staying within their advertised IP address boundaries) and exchanged upon a settlement-free basis (Ahmed, 2016). Adding the Transit Degree allowed us to gain a better understanding of the general upstream market importance of certain Autonomous Systems. Therefore, an Autonomous System in this section is labelled by reporting both the AS Number and its Transit Degree value derived from the CAIDA (2016a) AS-Rank dataset (ASN: Transit Degree).

For the purpose of exploring the centrality measurements in this section, we again placed the three Tamil Nadu mobile broadband operator networks into a two-dimensional Euclidean space to obtain the necessary graph visualisations. Here, we made use of the force-directed *Fruchterman-Reingold Layout*. We considered this layout valuable for the analysis since it does not apply any specific lengths to edge visualisations, resulting in vertices showing the seemingly same distances between each other. This allowed us to clearly identify those Autonomous System vertices of key interest (inhabiting certain centrality values). Moreover, colouring the key vertices of interest helped for identification purposes. We coloured vertices with low centrality value ‘red’ and vertices with high centrality ones ‘blue’. For comparability purposes of our obtained results, we set the following standards for the graph visualisation layout properties:

- Area of visualisation: 10000.0.
- Gravity attraction: 10.0.
- Edge thickness: $G_{Aircel_AS} : 0.25$, $G_{Bharti\ Airtel_AS} : 0.5$ and $G_{Vodafone_AS} : 0.05$.
- Graph structuring by *Degree Centrality*.

In the next paragraphs, we first visualise and compare the centrality metrics for G_{Aircel_AS} , followed by $G_{Bharti\ Airtel_AS}$ and lastly $G_{Vodafone_AS}$.

Aircel

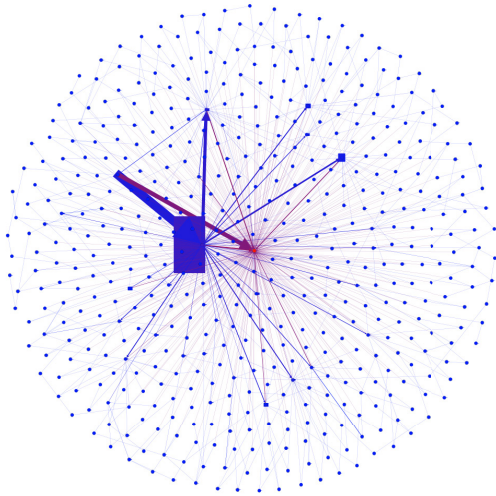
Looking at the *Degree Centrality* for G_{Aircel_AS} revealed seven Autonomous Systems, labelled as (ASN: Transit Degree), with a high *Degree Centrality*. The identified Autonomous Systems were, ordered-by their values, first an entry named #N/A, representing *traceroute* terminations that will be neglected, then Tata Communications (America) Inc. (AS6453:643), Level 3 Communications Inc. (AS3356:4160), PJSC Rostelcom (AS12389:632), Cogent Communications (AS174:4541), and NTT America Inc. (AS2914:1187). These vertices provide interesting insights. Based on our empirical evidence, merged with the information on AS names derived from Hurricane Electric (2016), we can see that G_{Aircel_AS} has a strong *Out-Degree* connectivity reliance on Tata Communications (America) Inc. (AS6453:643) as well as on Level 3 Communications Inc. (AS3356:4160). Information on the *In-Degrees* connectivity does not provide additional insights. Moreover, the *Degree Centrality* for G_{Aircel_AS} showed the presence of Autonomous Systems with a very high Transit Degrees. These high ranked Autonomous Systems in the CAIDA (2016a) dataset were Level 3 Communications Inc. (AS3356:4160) and Cogent Communications (AS174:4541).

The *Closeness Centrality*, showing how close a vertex is to the entire graph for G_{Aircel_AS} , identified the role of additional Autonomous Systems. In detail, the exploration revealed, ordered by the *Closeness Centrality* value, the following Autonomous Systems: Taiwan Fixed Network (AS9924:48), Nextweb Inc. (AS7829:21), Hong Kong Broadband Network Ltd. (AS9269:19), Singapore Telecommunications Ltd. (AS7473:261) and Tata Communications (formerly VSNL), (AS4755:400) as the centrally closest Autonomous Systems to the Aircel network. This is interesting since it shows that these Autonomous Systems are very well connected amongst the other Autonomous Systems in the network generated by the connectivity data of G_{Aircel_AS} . Moreover, none of them has very high Transit Degrees in the CAIDA (2016a) database, which potentially indicates possible presence of peering relationships between Autonomous Systems with a high Transit Degree.

The analysis of the *Betweenness Centrality* reveals the identity of those Autonomous System vertices that were key in ‘building bridges’ between other Autonomous Systems within the graph generated by the set of *traceroutes* for the Aircel network. These vertices are of key importance for the analysis of emerging structural internetworking bottlenecks

in the three mobile broadband operator networks. In descending order, the vertices with the highest *Betweenness Centralities* in G_{Aircel_AS} were Tata Communications (America) Inc. (AS6453:643), #N/A, Level 3 Communications Inc. (AS3356:4160), and PJSC Rostelcom (AS12389:632). This evidence confirms the previous findings about G_{Aircel_AS} relying on both Tata Communications (America) Inc. (AS6453:643) and Level 3 Communications Inc. (AS3356:4160) for its internetworking connectivity. Interestingly, PJSC Rostelcom (AS12389:632) is a Russian *Internet Service Provider*, which peers with both Tata Communications (America) Inc. (AS6453:643) and Level 3 Communications Inc. (AS3356:4160), see Hurricane Electric (2016).

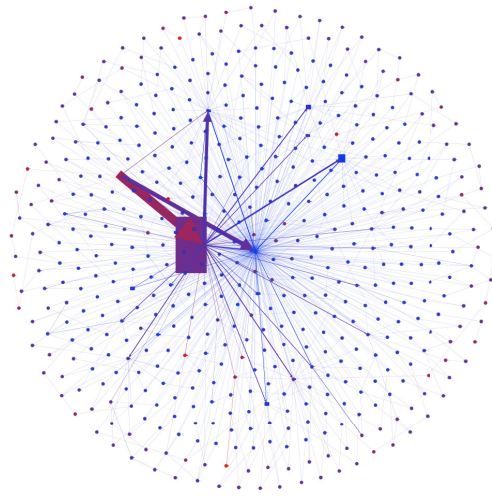
Lastly, the analysis of the *Eigenvector Centrality* in G_{Aircel_AS} revealed Tata Communications (America) Inc. (AS6453:643), followed by Cogent Communications (AS174:4541), Level 3 Communications Inc. (AS3356:4160) NTT America Inc. (AS2914:1187) above an *Eigenvector Centrality* threshold of 0.01. These Autonomous Systems hence show a strong internetworking influence within Aircel, representing the most valuable vertices to connect to, since high-scoring vertices contribute more to the influence than low-scoring ones. This indicates a structural reliance of G_{Aircel_AS} on these large Tier1 *Internet Service Providers*.



G_{Aircel_AS}

Degree Centrality

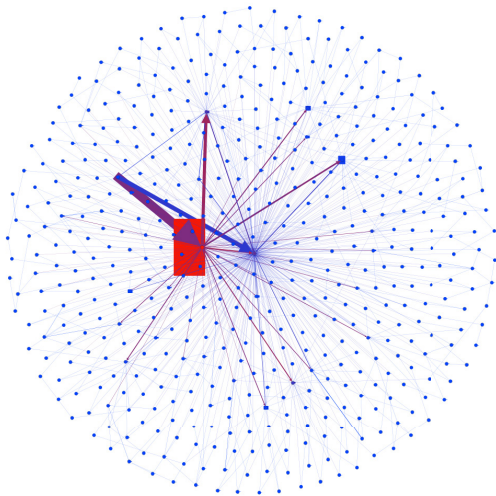
small = blue, high = red.



G_{Aircel_AS}

Closeness Centrality

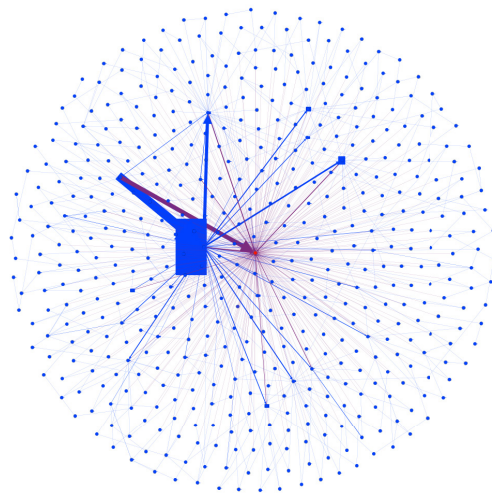
small = blue, high = red.



G_{Aircel_AS}

Betweenness Centrality

small = blue, high = red.



G_{Aircel_AS}

Eigenvector Centrality

small = blue, high = red.

Figure 4-15: Aircel graph visualisations centrality metrics using the Fruchterman-Reingold Layout at AS granularity.

Bharti Airtel

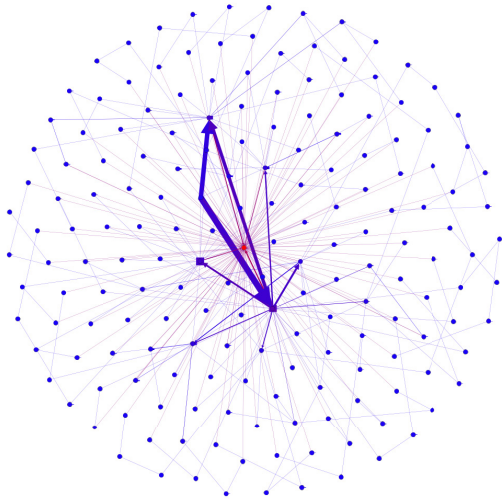
Starting again with the *Degree Centrality* for $G_{Bharti\ Airtel_AS}$, indicated three Autonomous Systems with the highest number of connections. Similar to G_{Aircel_AS} ,

Level 3 Communications Inc. (AS3356:4160) incorporated a strong *Degree Centrality*. Moreover, $G_{Bharti Airtel_AS}$ also showed Cogent Communications (AS174:4541) and Bharti Airtel Ltd. (AS9498:537) with the strongest *Out-Degree Centralities*. These three Autonomous System vertices again show mostly high *Out-Degree Centralities*, rather than *In-Degree Centralities*. The previous findings that $G_{Bharti Airtel_AS}$ strongly connects to itself is, therefore, supported.

The *Closeness Centrality* for $G_{Bharti Airtel_AS}$ exposed five vertices with the highest closeness to the entire Bharti Airtel network. These vertices were, in descending order: Telemar Norte Lesta S.A (AS7738:182), IP-Only Networks AB (AS12552:191), Brasil Telecom S/A (AS8167:202), Intelsat Global Service Corp. (AS22351:19) and finally COLT Technology Services Group Ltd (AS8820:655). Their importance represents properties of the most number of shortest paths between themselves and other Autonomous Systems in $G_{Bharti Airtel_AS}$. An Autonomous System could theoretically choose these vertices for establishing efficient connections.

By calculating the *Betweenness Centrality* for $G_{Bharti Airtel_AS}$, we detected Cogent Communications (AS174:4541), True International Gateway Co. (AS38082:115) and Bharti Airtel Ltd. (AS9498:537) had the highest *Betweenness Centralities*. Hence, these Autonomous Systems have the greatest number of shortest paths passing through them. Of great interest is the emerging evidence showing that $G_{Bharti Airtel_AS}$ relies on Cogent Communications (AS174:4541), an Autonomous System with a very well ranking in the CAIDA (2016a) AS-Rank.

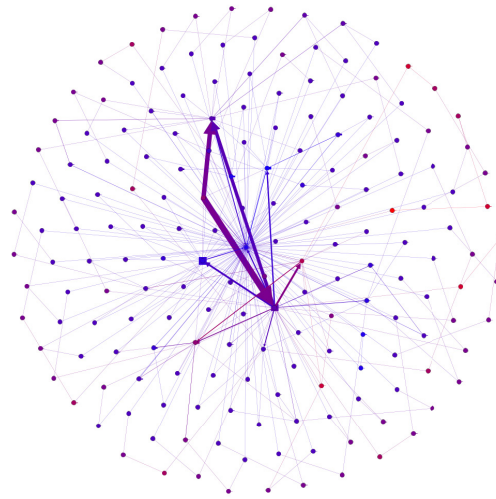
Lastly, those vertices that showed the highest *Eigenvector Centrality* for $G_{Bharti Airtel_AS}$, representing vertices with greatest connectivity influence within Bharti Airtel's network were, in descending order, Bharti Airtel Ltd. (AS9498:537), Level 3 Communications Inc. (AS3356:4160), Breeze Network (AS34661:5), Cogent Communications (AS174:4541), NTT America Inc. (AS2914:1187), the Amsterdam Internet Exchange (AS1200:0), Hurricane Electric Inc. (AS6939:3703) and Transtelecom (AS20485:1598). These reported Autonomous System vertices showed an *Eigenvector Centrality* over the threshold of a 0.15 value. These vertices, therefore, are those with highest influence, representing the most valuable Autonomous Systems to connect to for internetworking purposes (see the following Figure 4-16).



$G_{Bharti\ Airtel_AS}$

Degree Centrality

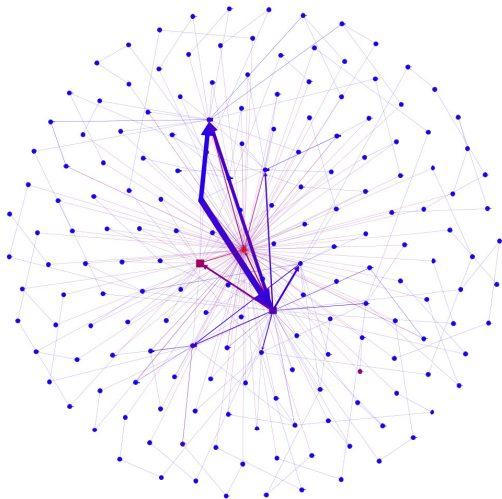
small = green, high = blue.



$G_{Bharti\ Airtel_AS}$

Closeness Centrality

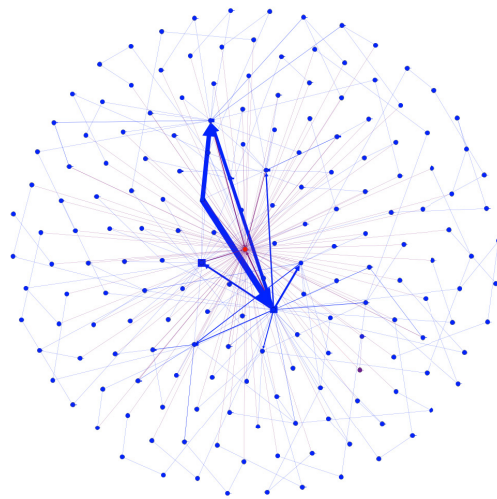
small = green, high = blue.



$G_{Bharti\ Airtel_AS}$

Betweenness Centrality

small = red, high = blue.



$G_{Bharti\ Airtel_AS}$

Eigenvector Centrality

small = red, high = blue.

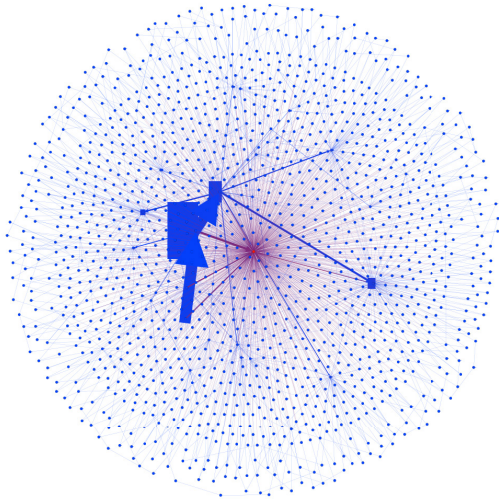
Figure 4-16: Bharti Airtel graph visualisations centrality metrics using the Fruchterman-Reingold Layout at AS granularity.

Vodafone

The *Degree Centrality* for $G_{Vodafone_AS}$ indicated four Autonomous System vertices with strong *Degree Centralities*. In descending order, these vertices were Cogent Communications (AS174:4541), Telia Company AB (AS1299:1010), Cable and Wireless Worldwide plc. (AS1273:355) and Tinet SpA (AS3257:1085). Moreover, the graph visualisation revealed some other Autonomous System vertices also with a relatively high *Degree Centrality*. These vertices included Internet Service Providers such as Level 3 Communications Inc. (AS3356:4160) and NTT America Inc. (AS2914:1187).

The Autonomous Systems vertices with the highest *Closeness Centrality* in $G_{Vodafone_AS}$ were Blix Solutions AS (AS50304:563), Etisalat (AS8966:140), Tecnocratica Centro de Datos (AS15954:7), and the Belarusian Cloud Technologies JLLC (AS60330:11). Again, the network showed numerous other Autonomous Systems with a strong *Closeness Centrality*.

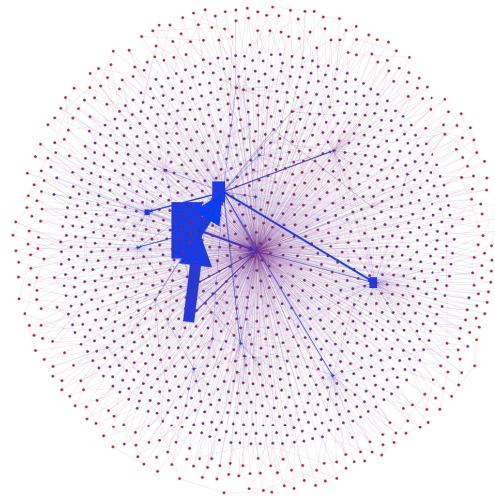
Moreover, $G_{Vodafone_AS}$ revealed mainly two vertices with a strong *Betweenness Centrality* and hence the greatest number of shortest paths passing between them. These Autonomous Systems were Telecom Italia Sparkle SpA (AS6762:351) and Level 3 Communications Inc. (AS3356:4160). This indicates the important role played by these two ASes in establishing connections to the periphery of $G_{Vodafone_AS}$. Due to their unavoidability, under given interconnection policies, both of these Autonomous Systems show the potential for exerting a strong bargaining position and hence market power towards $G_{Vodafone_AS}$. Again, Level 3 Communications Inc. (AS3356:4160) also shows a very high Transit Degree compared to Telecom Italia Sparkle SpA (AS6762:351). This shows that Telecom Italia Sparkle SpA (AS6762:351) is likely to align with more peering relationships, rather than transit ones. It seems that $G_{Vodafone_AS}$ connects with these Autonomous Systems independent of the given *traceroute* destinations. Those vertices that showed the highest *Eigenvector Centrality* for $G_{Vodafone_AS}$, representing vertices with greatest connectivity influence within Vodafone's network, were, again in descending order, Vodafone India Ltd. (AS55410:157) with the maximum *Eigenvector Centrality* 1.0, followed by Bharti Airtel Ltd. (AS9498:537), Telstra Global (AS4637:226) and Vodafone's Cable and Wireless Worldwide plc. (AS1273:355) subsidiary, all ranging above an *Eigenvector Centrality* threshold of 0.0010. These valuable Autonomous System vertices represent again those with the highest internetworking influence in the network (see Figure 4-17 on the next page).



$G_{Vodafone_AS}$

Degree Centrality

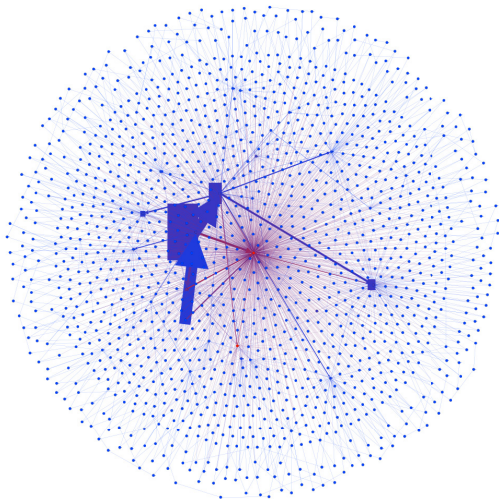
small = red, high = blue.



$G_{Vodafone_AS}$

Closeness Centrality

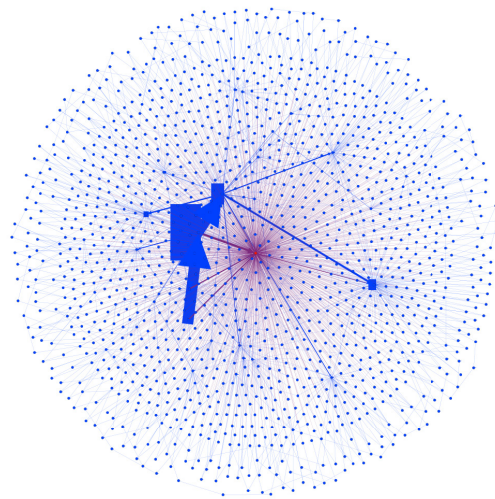
small = red, high = blue.



$G_{Vodafone_AS}$

Betweenness Centrality

small = red, high = blue.



$G_{Vodafone_AS}$

Eigenvector Centrality

small = red, high = blue.

Figure 4-17: Vodafone graph visualisations centrality measurements using the Fruchterman-Reingold Layout at AS granularity.

Summarising, the Centrality Analysis for the three providers indicated structurally important Autonomous Systems. However, we note that it's hard to obtain the right vertices by choosing them manually based on the graph visualisation. While our findings seem convincing, we provide them with the following limitations. Given the metric's theoretical grounding, we indicate that most metrics are unable to capture the properties of interest. The *Degree Centrality* does not take, for example, vertex importance as well as edge directions into consideration. Hence, this metric does not suit our directed network but it shows those vertices being important in the *power-law degree distribution*. Moreover, the *Closeness Centrality* and *Betweenness Centrality* both assume that communication within a network always follows the shortest paths. This is unlikely to be the case in a real-world network. Here, connectivity of *traceroutes* would not always follow the shortest paths to reach a destination. Based on these metrics alone, it seems difficult to indicate those Autonomous Systems with central properties regarding their connectivity importance. To find those Autonomous Systems, the following section covers the respective *k-core decomposition* algorithm of Alvarez-Hamelin et al. (2005b; 2008). We aim to detect those Autonomous Systems that were building very dense connections in the core of the operator networks, a critical indicator for the hierarchical organisation of the operators.

k-core decomposition (AS)

In this section, we will use the *k-core decomposition* spectral analysis to identify the most densely connected Autonomous System vertices for each of the graphs generated for the three Tamil Nadu mobile broadband operator' networks at AS granularity. Alvarez-Hamelin et al. (2008, p.390) argue that the *k-core decomposition* can be used to compare different granularities of the Internet structure for the purpose of revealing structural properties, as present in our work. Referring to the work of Alvarez-Hamelin et al. (2005b), this *k-core decomposition* reveals the specific roles and relevance of the vertices located in both, the network periphery and the network core. This method is frequently used for the analysis of Internet structures and employed by researchers at CAIDA (2015). Using a *k-core decomposition* algorithm allows for the division of graph visualisation into densely connected network subsets, called *k-cores*. These *k-cores* represent connectedness properties for the Autonomous System vertices in a given network, where a higher *k-core* indicates a set of more densely connected Autonomous Systems. These most densely-connected Autonomous System vertices in the mobile broadband operator network core provide both connectivity features amongst themselves and between this

central core and those Autonomous Systems located in the overall network periphery. Given our identified *k-cores*, this method allows for a clear identification and visualisation of some key hierarchical network properties.

Aircel

The *k-core decomposition* for G_{Aircel_AS} indicated 16 Autonomous Systems with the same high *k-core* value of ‘5’ as Table 4-15 below indicates. Besides being densely connected amongst each other, these Autonomous Systems also connect to the less-connected Autonomous Systems in the network periphery. Hence, the densest sub-graph represents the set of most influential Autonomous Systems for G_{Aircel_AS} . Therefore, these Autonomous Systems present the most robust routing capabilities. Moreover, this uncovers hierarchical properties of G_{Aircel_AS} . Interestingly, some of the Autonomous Systems in the densest sub-graph were previously revealed as being relevant regarding their *Out-Degree* features. These ASes were Tata Communications (America) Inc. (AS6453:643), Level 3 Communications Inc. (AS3356:4160), NTT America Inc. (AS2914:1187) and Cogent Communications (AS174:4541). Moreover, they all show a very high Transit Degree as well.

Aircel k-core vertices at AS granularity		
k-cores	ASN:Transitdegree (CAIDA, 2016a)	AS- Name (Hurricane Electric, 2016)
5	AS6453:643	Tata Communications (America) Inc.
5	AS4755:400	Tata Communications, formerly VSNL
5	AS174:4541	Cogent Communications
5	AS2914:1187	NTT America Inc.
5	AS3356:4160	Level 3 Communications Inc.
5	AS1299:1010	Telia Company AB
5	AS1239:667	Sprint
5	AS12552:191	IP-Only Networks AB
5	AS9002:1549	RETN Ltd.
5	AS1273:355	Cable and Wireless Worldwide plc
5	AS4323:2006	TW Telecom Holdings Inc.
5	AS3491:540	PCCW Global
5	AS4766:449	Korea Telecom
5	AS20485:1598	Closed Joint Stock Company TransTeleCom
5	AS3741:143	IS
5	AS7029:483	Windstream Communications Inc.
<p>Key</p> <p>AS: Autonomous System.</p> <p>ASN: Autonomous System Number.</p> <p>#N/A: Drop or termination of a <i>traceroute</i>.</p>		

Table 4-15: Aircel highest k-core vertices at AS granularity.

Additionally, none of the other Autonomous Systems in Table 4-15 above were previously indicated as relevant Autonomous Systems in the G_{Aircel_AS} network. We show that the *k-core decomposition* is a valuable addition to an in-depth analysis of operator networks at Autonomous System granularity. In comparison to the findings of G_{Aircel} 's *k-core decomposition* at IP granularity, the analysis at AS granularity uncovered

a considerably more detailed picture. Moreover, this analysis showed that G_{Aircel} seems to rely on a number of AS vertices with incredible high transit degrees. Figure 4-18 below visualises the *k-core decomposition* of G_{Aircel_AS} .

G_{Aircel_AS}

Highest k-core : 5.

Highest Autonomous System vertices, visualised in magenta in the centre of the graph visualisation (grey edges), see Table 4-15 above.

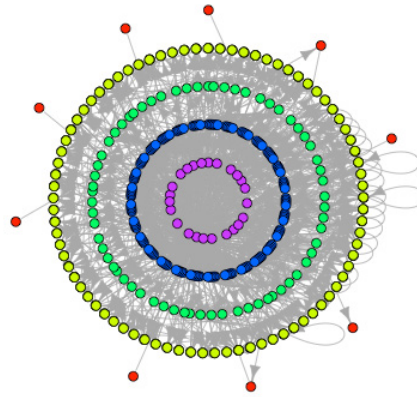


Figure 4-18: Aircel graph visualisation *k-core decomposition* at AS granularity.

Bharti Airtel

The *k-core decomposition* for $G_{Bharti\ Airtel_AS}$ indicated 5 Autonomous Systems that inhabited the same highest *k-core* value of 5, see Table 4-16 and Figure 4-19.

Compared to the previous centrality measurement analysis, the *k-core decomposition* of $G_{Bharti\ Airtel_AS}$ also revealed additional Autonomous System vertices that were of hierarchical importance to the network. While the *Degree Centrality* indicated the Out-Degree relevance of three Autonomous Systems, namely Level 3 Communications Inc (AS3356:4160), Cogent Communications (AS174:4541) and Bharti Airtel Ltd. (AS45609:3), the densest hierarchical layer of the *k-core decomposition* exposed $G_{Bharti\ Airtel_AS}$'s reliance on NTT America Inc. (AS2914:1187), Telia Company AB (AS1299:1010), and another Bharti Airtel Ltd. Autonomous System (AS9498:537).

Bharti Airtel k-core vertices at AS granularity		
k-cores	ASN:Transitdegree (CAIDA, 2016a)	AS- Name (Hurricane Electric, 2016)
5	AS9498:537	Bharti Airtel Ltd.
5	AS45609:3	Bharti Airtel Ltd.
5	AS174:4541	Cogent Communications
5	AS3356:4160	Level 3 Communications Inc.
5	AS2914:1187	NTT America Inc.
5	AS1299:1010	Telia Company AB
Key AS: Autonomous System. ASN: Autonomous System Number.		

Table 4-16: Bharti Airtel highest k-core vertices at AS granularity.

Interestingly, the *k-core decomposition* for $G_{Bharti\ Airtel}$ at IP granularity also revealed the presence of the two Bharti Airtel Ltd. (AS45609:3 and AS9498:537) Autonomous Systems and Level 3 Communications Inc. (AS336:4160) but failed to identify the other three Autonomous Systems shown at IP granularity. Since no other graph visualisation revealed these Autonomous Systems as being important hierarchical vertices, we show the importance of utilising different methods of analysis.

Furthermore, $G_{Bharti\ Airtel_AS}$ seem to also strongly rely on those Autonomous System vertices with high Transit Degrees. Hence, these vertices are of strong importance to any AS for reaching the periphery of the Internet, given their high appointing of transit relationships.

$G_{Bharti\ Airtel_AS}$

Highest k -core : 5

Highest Autonomous System vertices, visualised in magenta in the centre of the graph visualisation (grey edges), see Table 4-16 above.

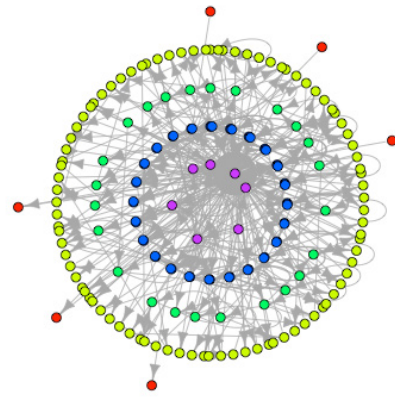


Figure 4-19: Bharti Airtel graph visualisation k -core decomposition at AS granularity.

Vodafone

Lastly the k -core decomposition of $G_{Vodafone_AS}$ shows one Autonomous System in the densest k -core 10 being #N/A, followed by 8 Autonomous Systems in the second highest k -core of 8. These Autonomous Systems are visible in the following Table 4-17.

We have already identified some of these AS vertices in the previous analysis of $G_{Vodafone_AS}$. This was namely Level 3 Communications Inc. (AS3356:4160) in the analysis of the *Betweenness Centrality*, and Cogent Communications (AS174:4541) as well as Cable and Wireless Worldwide plc. (AS1273:355) when studying the network's *Degree Centrality*. However, the *Degree Centrality* for $G_{Vodafone_AS}$ also identified other Autonomous System vertices that were not present in the k -core decomposition's densest k -core sub-graph, Telia Company AB (AS1299:1010) and Tinet SpA (AS3257:1085). This indicates that Telia Company AB and Tinet SpA, while showing high *Out-Degrees*, are not necessarily of strong hierarchical relevance to $G_{Vodafone_AS}$.

Vodafone k-core vertices – AS granularity		
k-cores	ASN:Transitdegree (CAIDA, 2016a)	AS- Name (Hurricane Electric, 2016)
10	#N/A	#N/A
8	AS55410:157	Vodafone India Ltd.
8	AS1273:355	Cable and Wireless Worldwide plc.
8	AS4538:21	China Education and Research Network Center
8	AS174:4541	Cogent Communications
8	AS3356:4160	Level 3 Communications Inc.
8	AS2914:1187	NTT America Inc.
8	AS3491:540	PCCW Global
8	AS7018:2320	AT&T Services Inc.
Key AS: Autonomous System. ASN: Autonomous System Number. #N/A: Drop or termination of a <i>traceroute</i> .		

Table 4-17: Vodafone highest k-core vertices at AS granularity.

Moreover, this analysis detected more Autonomous System vertices than the *k-core decomposition* at IP granularity. At IP granularity, the *k-core decomposition* indicated a hierarchical reliance on Vodafone India Ltd. (AS55410:157), the China Education and Research Network Center (AS4538:21) and Cable and Wireless Worldwide plc. (AS1273:355). Therefore, the *k-core decomposition* at AS granularity enhanced the overall understanding of the role played by some of the most important Autonomous Systems. Interestingly, $G_{Vodafone_AS}$ also shows two Autonomous Systems with very high Transit Degrees, namely Cogent Communications (AS174:4541) and Level 3 Communications Inc. (AS3356:4160). However, all of the Autonomous Systems in *k-core* 8 strongly participate in the Internet traffic transmission originated from the issued Vodafone's SIM card.

$G_{Vodafone_AS}$

Highest k-core: 10.

Highest Autonomous System vertices, visualised in red and dark blue in the centre of the graph visualisation (grey edges), see Table 4-17 above.

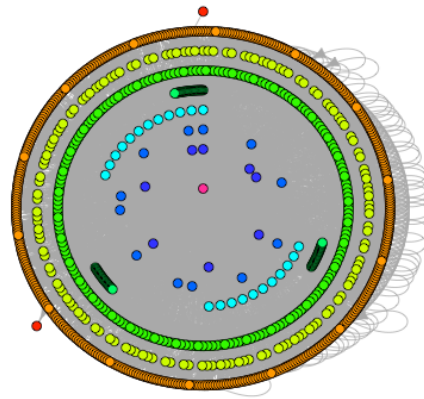


Figure 4-20: Vodafone graph visualisation *k*-core decomposition at AS granularity.

Summarising, the *k*-core decomposition at AS granularity revealed the most strongly hierarchical Autonomous System vertices for each of the three operator networks. Moreover, the detection of these key vertices was more apparent than when using the *k*-core decomposition at IP granularity. Overall, the *k*-core decomposition can be considered a very useful method for detecting the densely connected Autonomous System vertices and therefore important hierarchical markers.

Summary Graph Visualisation Analysis (AS)

Overall, we value the exploratory approach of utilising different graph visualisation analysis. Firstly, this section showed that none of the three operator networks employed *Small-World Network* properties. This was not surprising given the same indication at IP granularity. Next, we showed that the three operator networks are very well modelled using the *Scale-Free* Barabási-Albert Model B. The analysis of Centrality Metrics then revealed a number of Autonomous Systems with varying importance to the three operator networks. However, we also identified the limits of these measurements to study *traceroute* hop observations. Given these limitations, e.g. that the *Closeness*- and *Betweenness Centralities* both assume that connections are following the shortest paths, we indicate that the *Eigenvector Centrality* might be an interesting new metric for studying Autonomous System influence and hence hierarchical upstream Internet market structures (see also section 2.3.3). We then analysed the most densely connected Autonomous Systems in the core of the operator networks by using the *k*-core decomposition. Here, those Autonomous System vertices with the highest *k*-core indicate

signs of hierarchical organisation and were most clearly visible in the graph visualisations of Aircel and Vodafone. These vertices were considered as the emerging key players in determining the network's hierarchical structure, based on their connectivity importance.

4.3.8 Summary Complex Network Analysis (AS)

Summarising the *Complex Network Analysis* at AS granularity, we indicated the main differences from the mapping of the previous IP granularity. Firstly, the General Network metric analysis provided us with indications of influential Autonomous Systems that were more often traversed than other Autonomous Systems in the operator networks. Moreover, we indicated great *Clustering Coefficient* differences in the three operator networks, pointing towards different hierarchical upstream Internet market structures. Next, we suggested that the three operator networks are likely to follow *Scale-Free Network* models and tested this by using the *Graph Visualisation Analysis*. Moreover, this *Graph Visualisation Analysis* exposed some interesting Autonomous Systems. We showed that only the *Eigenvector Centrality* might be valuable for analysing active Internet periphery measurements of *traceroutes*. Moreover, the *k-core decomposition* in particular helped us to gain a deeper understanding of the most densely connected Autonomous Systems within the operator networks. These ASes seem to play a crucial role in the upstream connectivity given their great linkages to other influential Autonomous Systems. Hence, we consider those Autonomous Systems as key players for determining the operators' upstream Internet market structures.

Hence, the following section aims to better understand the economic relationships between the identified key Autonomous Systems for the three Tamil Nadu operators. This should provide an in-depth understanding of the underlying economics of network interactions, both within the densest sub-graph, as well as to the less-densely connected Network periphery.

4.4 Autonomous System Relationships

In this section, we will elaborate a first economic intuition of the internetworking relationships between Autonomous System. To achieve this, we will integrate additional information on the key relationships between the Autonomous Systems of relevance for the three Tamil Nadu mobile broadband operator networks. Following, we are fusing our primary dataset with three additional secondary ones:

- i. The CAIDA (2016a) AS-Rank dataset, derived from CAIDA's Archipelago Measurement Infrastructure.
- ii. The CAIDA (2016b) AS-Relationship dataset, which covers the inferred Customer Cones from publicly available Border Gateway Protocol (BGP) data.
- iii. The Hurricane Electric (2016) BGP Routing Tables, covering the IPv4 Route Propagation graphs for Autonomous System Numbers.

For each of the three Tamil Nadu mobile broadband operator networks, we perform an analysis of the AS relationships based on the consideration of two distinct Autonomous System properties. First, whether an Autonomous Systems belongs to the central network core, as calculated using the *k-core decomposition* explained above. Second, for any pair of directly connected Autonomous Systems, we will consider the nature of their business relationship, as inferred from the CAIDA (2016b) AS-Relationship dataset.

Here, the CAIDA (2016a) AS-Rank data includes information on the Customer Cone (see also section 2.1.2) of a particular Autonomous System. According to Luckie et al. (2013), the Customer Cone represents the set of Autonomous Systems that one AS may reach by recursively following its customer links.

Next, we analyse the strongest relationships between all Autonomous Systems in the three networks, as derived from the *k-core decomposition* above. We merge these key relationships with the CAIDA (2016b) AS-Relationship dataset to explore their inferred economic nature. Given the evidence, we then visualise the mobile broadband operator network graphs focussing on the identified economic relationships amongst the Autonomous Systems. This reveals some interesting economic differences between our three Tamil Nadu mobile broadband operator networks.

4.4.1 Aircel Autonomous System Relationships

Table 4-18 below lists the Autonomous Systems belonging to the highest *k-core* of

G_{Aircel_AS} , as derived from the *k-core decomposition* at AS granularity, alongside their associated Customer Cone Sizes (CCS) from the CAIDA (2016a) AS-Rank. These findings indicate that the highest *k-core* in the *k-core decomposition* of G_{Aircel_AS} includes a multitude of Autonomous Systems with a high Customer Cone Size. Interestingly, and in descending order, these ASes represent the top 4 Autonomous Systems in the CAIDA (2016a) Customer Cone ranking.

The large number of Autonomous Systems in the highest *k-core* of G_{Aircel_AS} shows that Aircel makes use of multiple AS relationships, highly Customer Cones ranked, for its complete internetworking activities.

However, the following Table 4-18 does not show the entire picture of the complex routing situation. To discover more information, we utilise the BGP Route Propagations graphs for Aircel using Hurricane Electric (2016). These route propagation graphs revealed that G_{Aircel_AS} only shows two direct links to other Autonomous Systems, namely Bharti Airtel Ltd. (AS9498:537) and Dishnet Wireless Ltd. (AS55713:4), (see Appendices). Bharti Airtel Ltd. has, as seen from the secondary CAIDA (2016b) AS-Relationship dataset and Hurricane Electric (2016), only one direct link to Tata Communications (formerly VSNL), (AS4755:400) among the ASes mentioned above.

Looking at the BGP Route Propagation graphs for the Autonomous System of Aircel in Hurricane Electric (2016), one can see that its upstream connectivity is not only reliant on Bharti Airtel Ltd. (AS9498:537) but also on Tata Communications (formerly VSNL), (AS4755:400). Once a connection reaches this vertex, it links to Tata Communications (America) Inc. (AS6453:643). Here, it finally reaches some of the other large Autonomous Systems mentioned in Table 4-18, (see Appendices). Furthermore, according to the CAIDA (2016b) AS relationship dataset, it is worth to notice that only one of the Aircel Autonomous System relationships between those Autonomous Systems identified in the highest *k-core*, is of a peering nature. This relationship is the one existing between Tata Communications (America) Inc. (AS6453:643) and Level 3 Communications Inc. (AS3356:4160), see column 4 row 8 in Table 4-19.

Aircel highest k-core Autonomous Systems ranked by Customer Cone Size		
ASN: Transit Degree (CAIDA, 2016a)	AS- Name (Hurricane Electric, 2016)	CCS (CAIDA, 2016a)
AS3356:4160	Level 3 Communications Inc.	29,494
AS174:4541	Cogent Communications	23,299
AS3356:4160	Level 3 Communications Inc.	29,494
AS1299:1010	Telia Company AB	21,954
AS2914:1187	NTT America Inc.	18,991
AS6453:643	Tata Communications (America) Inc.	12,300
AS1273:355	Cable and Wireless Worldwide plc	5,878
AS9002:1549	RETN Ltd.	3,820
AS3491:540	PCCW Global	3,572
AS20485:1598	Closed Joint Stock Company TransTeleCom	3,447
AS1239:667	Sprint	3,439
AS4323:2006	TW Telecom Holdings Inc.	2,184
AS4766:449	Korea Telecom	959
AS4755:400	Tata Communications, formerly VSNL	732
AS12552:191	IP-Only Networks AB	219
AS3741:143	IS	148
<p>Key</p> <p>AS: Autonomous System.</p> <p>ASN: Autonomous System Number.</p> <p>CCS: Customer Cone Size.</p>		

Table 4-18: Aircel customer cones.

The analysis of the merged datasets, considering the

- i. Autonomous Systems belonging to the highest core calculated through the *k-core decomposition* and the edge weights (both obtained from the primary collected data using Portolan (2016)) and further elaborated by the author
- ii. CAIDA (2016a) Customer Cone rank
- iii. inferred business relationships of the CAIDA (2016b) AS – Relationships dataset

, shows interesting relevant features for G_{Aircel_AS} 's upstream Internet market structure. This reveals new AS relationships from an actively measured Internet periphery perspective that do not appear as direct business relationships in the CAIDA (2016b) AS – Relationships dataset.

Aircel AS Relationships using CAIDA (2016a)			
SourceASN : Transit Degree, (AS Rank)	TargetASN : Transit Degree, (AS Rank)	Edge Weight	CAIDA AS Relationship (CAIDA, 2016b)
AS6453:643, (6)	AS6453:643, (6)	1,367	#N/A
AS4755:400, (61)	AS6453:643, (6)	212	#N/A
AS174:4541, (2)	AS174:4541, (2)	194	#N/A
AS2914:1187, (4)	AS2914:1187, (4)	107	#N/A
AS4755:400, (61)	#N/A	105	#N/A
AS6461:1381, (16)	AS6461:1381, (16)	82	#N/A
AS7922:172, (27)	AS7922:172, (27)	77	#N/A
AS6453:643, (6)	AS3356:4160, (1)	76	p2p
AS3356:4160, (1)	AS3356:4150, (1)	64	#N/A
AS1299:1010, (3)	AS1299:1010, (3)	58	#N/A
Key AS: Autonomous System. ASN: Autonomous System Number. CCS: Customer Cone Size.			

Table 4-19: Aircel AS Relationships.

Merging the CAIDA (2016b) AS relationship dataset with the collected Aircel internetworking data, we can see that 2.17% of the collected AS relationships were of *customer-to-provider* nature. Additionally, 3.33% of the Autonomous System relationships represented *peer-to-peer* relationships and 3.94% *provider-to-customer* ones. Moreover, 61.09% of all AS relationships represented AS-internal connections (*traceroute* hops from the dataset generated from the Aircel operator directly linking a source with a target IP addresses, both belonging to the same Autonomous System), and 13.18% of all AS relationships resulted from *traceroute* terminations. Finally, 16.30% of all observed Autonomous System relationships do not appear as business relationships in the CAIDA (2016b) AS-Relationship dataset (see Table 4-20).

Aircel AS Relationships Overview using CAIDA (2016b)				
Type of AS Relationship (CAIDA, 2016b)	Description	Number of Edge observations	Edge Weights (Traversals)	In percentage of total edge weights
-1	c2p relationship	12	103	2.17%
0	p2p relationship	19	158	3.33%
1	p2c relationship	83	187	3.94%
#N/A	CAIDA (2016b) undetected	428	774	16.30%
#N/A	AS-internal edge	237	2,901	61.09%
#N/A	<i>Traceroute</i> Termination	362	626	13.18%
Key AS: Autonomous System. c2p: Customer-to-provider relationship. p2c: Provider-to-customer relationship. p2p: Peer-to-peer Relationship. #N/A: No relationship in CAIDA (2016b) available.				

Table 4-20: Aircel AS Relationships overview.

Plotting the merged data into a two-dimensional Euclidean Space in Figure 4-21 below reveals some interesting features. Based on these graph visualisations, we note that Aircel

is seemingly making more use of *provider-to-customer* (p2c) relationships than both *peer-to-peer* (p2p), or *customer-to-provider* (c2p) ones. Moreover, we clearly reveal those AS relationships with the highest edge weights given their different edge colourings. These visualisations are of particular interest and value as it indicates the economic nature of the key relationships within the network of G_{Aircel_AS} .

G_{Aircel_AS}

Edge thickness: 0.5

Red edges: p2c link.

Green edges: p2p link.

Blue edges: c2p link.

Yellow edges: #N/A
(link not available)

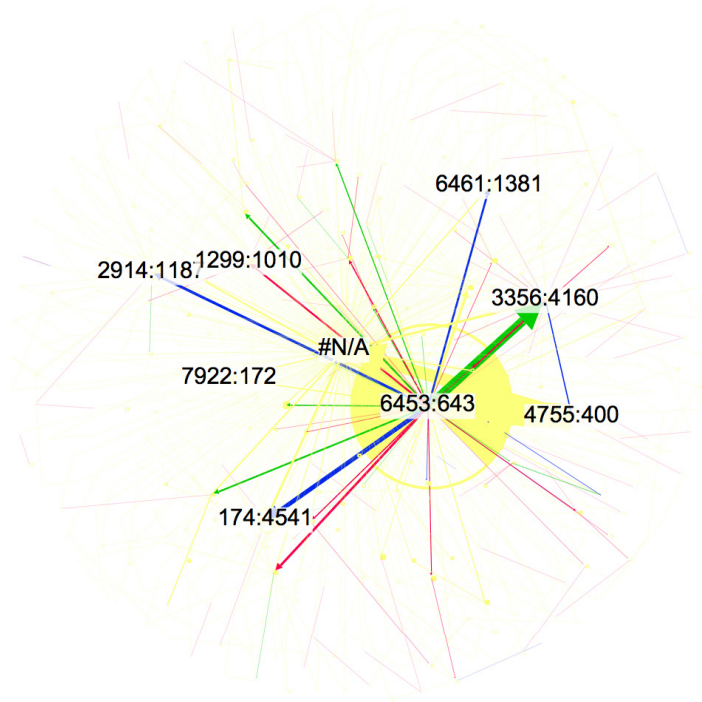


Figure 4-21: Aircel graph visualisation with relationship colouring.

Summarising, G_{Aircel_AS} shows a reliance on the Indian ASes of Bharti Airtel Ltd. and Tata Communications (formerly VSNL), providing additional indicators of a hierarchical ordering. None of the crucial relationships (highest edge weights) in G_{Aircel_AS} were *peer-to-peer* ones. This indicates the presence of higher upstream connectivity costs. Considering routing and structuring, our analysis revealed some hidden features of Aircel's operator network. These features were not observable through the previously available datasets.

4.4.2 Bharti Airtel Autonomous System Relationships

By following the same approach as above, Table 4-21 below lists the Autonomous Systems belonging to the highest *k-core* of $G_{Bharti_Airtel_AS}$, derived from the *k-core decomposition* at AS granularity, alongside their associated Customer Cone Sizes (CCS).

Interestingly, $G_{Bharti\ Airtel_AS}$ shows the same four Tier-1 *Internet Service Providers* (with a high Customer Cone Size) than G_{Airtel_AS} . The ASes in the highest k -core of $G_{Bharti\ Airtel_AS}$ indicates that Bharti Airtel makes use of a few Autonomous System relationships with a high Customer Cones ranking for its complete internetworking activities.

Bharti Airtel highest k-core Autonomous Systems ranked by Customer Cone Size		
ASN : Transit Degree (CAIDA, 2016a)	AS- Name (Hurricane Electric, 2016)	CCS (CAIDA, 2016a)
AS3356:4160	Level 3 Communications Inc.	29,494
AS174:4541	Cogent Communications	23,299
AS1299:1010	Telia Company AB	21,954
AS2914:1187	NTT America Inc.	18,991
AS9498:537	Bharti Airtel Ltd.	1,475
AS45609:3	Bharti Airtel Ltd.	1
Key AS: Autonomous System. ASN: Autonomous System Number. CCS: Customer Cone Size.		

Table 4-21: Bharti Airtel customer cone.

Looking at the BGP Route Propagation graphs for Bharti Airtel in Hurricane Electric (2016), one can see that its upstream connectivity is reliant on their own directly reachable Autonomous Systems (AS9498 and AS45609). Once a connection reaches this vertex, it links to some of the large Tier-1 Autonomous Systems mentioned in Table 4-21 above, (see Appendices in Chapter 9).

Moreover, according to the CAIDA (2016b) AS relationship dataset, it is worth to notice that none one of the Bharti Airtel Autonomous System relationships with the strongest edge weight was identified (see Table 4-22 below). Additionally, the strong relationship between Bharti Airtel Ltd. (AS45609:3) and Level 3 Communications Inc. (AS3356:4160) in row 4 of Table 4-22, as measured through the edge weight, is not mentioned in the BGP Route Propagation graph of Hurricane Electric (2016), (see

Appendices).

Bharti Airtel AS Relationships using CAIDA (2016a)			
SourceASN : Transit Degree, (AS Rank)	TargetASN : Transit Degree, (AS Rank)	Edge Weight	AS Relationship (CAIDA, 2016b)
AS174:4541, (2)	AS174:4541, (2)	69	#N/A
AS9498:537, (33)	AS9498:537, (33)	61	#N/A
AS45609:3, (7469)	AS9498:537, (33)	55	#N/A
AS45609:3, (7469)	AS3356:4160, (1)	40	#N/A
AS9498:537, (33)	AS3356:4160, (1)	30	#N/A
AS3356:4160, (1)	AS3356:4160, (1)	30	#N/A
AS6939:3703	AS6939:3703, (8)	26	#N/A
AS2914:1187, (4)	AS2914:1187, (4)	23	#N/A
AS9498:537, (33)	AS34661:5, (3327)	17	#N/A
AS9498:537, (33)	AS174:4541, (2)	16	#N/A
Key AS: Autonomous System. ASN: Autonomous System Number. CCS: Customer Cone Size.			

Table 4-22: *Bharti Airtel AS Relationships.*

Merging the collected Bharti Airtel internetworking data with the CAIDA (2016b) AS relationship dataset, we can see that 0.42% of the collected AS relationships were of *customer-to-provider* nature. Additionally, 2.51% of the AS relationships represented *peer-to-peer* relationships and 2.30% *provider-to-customer* ones. Moreover, 36.19% of all AS relationships represented AS-internal connections (*traceroute* hops from the dataset generated from the Airtel operator directly linking a source with a target IP addresses, both belonging to the same Autonomous System), and 15.69% of all AS relationships resulted from *traceroute* terminations. Finally, a high 36.19% of all observed Autonomous System relationships do not appear as business relationships in the

CAIDA (2016b) AS Relationship dataset (see Table 4-23 below).

Bharti Airtel AS Relationships Overview using CAIDA (2016b)				
Type of AS Relationship (CAIDA, 2016b)	Description	Number of Edge observations	Edge Weights (Traversals)	In percentage of total edge weights
-1	cp2 relationship	4	4	0.42%
0	p2p relationship	15	24	2.51%
1	p2c relationship	19	22	2.30%
#N/A	CAIDA (2016b) undetected	151	346	36.19%
#N/A	AS-internal edge	82	410	42.89%
#N/A	<i>Traceroute</i> Termination	116	150	15.69%
<p>Key</p> <p>AS: Autonomous System.</p> <p>c2p: Customer-to-provider relationship.</p> <p>p2c: Provider-to-customer relationship.</p> <p>p2p: Peer-to-peer Relationship.</p> <p>#N/A: No relationship in CAIDA (2016b) available.</p>				

Table 4-23: Bharti Airtel AS Relationships overview.

Just like before, plotting the merged data into a two-dimensional Euclidean Space in Figure 4-22 below reveals that Bharti Airtel is making most use of *peer-to-peer* relationships. Again, we clearly reveal those AS relationships with the highest edge weights given their different edge colourings. Compared to G_{Airtel_AS} , the graph visualisation of $G_{Bharti\ Airtel_AS}$ seems more balanced while showing a greater presence of (settlement-free) *peer-to-peer* relationships, indicating a potential connectivity advantage over G_{Airtel_AS} .

$G_{Bharti\ Airtel_AS}$

Edge thickness: 1.0

Red edges: p2c link.

Green edges: p2p link.

Blue edges: c2p link.

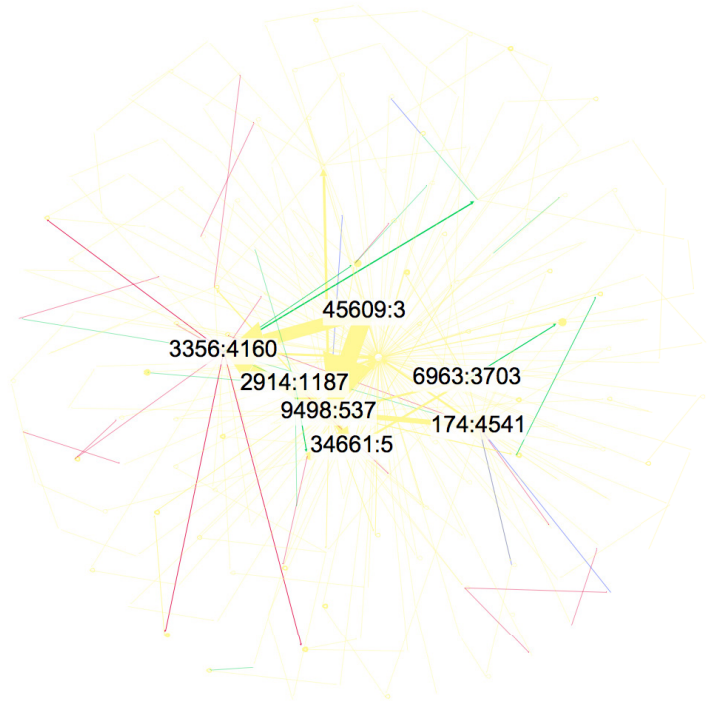
Yellow edges: #N/A
(link not available).

Figure 4-22: Bharti Airtel graph visualisation with relationship colouring.

Summarising, the evidence showed that the upstream connectivity in $G_{Bharti\ Airtel_AS}$ is strongly reliant on their Bharti Airtel Ltd. (AS9498:537) Autonomous System. Unlike the BGP Route Propagation graphs in Hurricane Electric (2016), the analysis exposed a direct connection to Level 3 Communications Inc. (AS3356:4160). This is particularly interesting since Level 3 Communications Inc. (AS3356:4160) represents a major Tier-1 *Internet Service Provider*. Just like before, this analysis revealed some hidden routing and structuring features of Bharti Airtel's operator network. These features were not observable through the previously available datasets.

4.4.3 Vodafone Autonomous System Relationships

By following the same approach as for Airtel and Bharti Airtel above, Table 4-24 below lists the Autonomous Systems belonging to the highest k-core of $G_{Vodafone_AS}$, as derived from the k-core decomposition at AS granularity, alongside their associated Customer Cone Sizes (CCS). Interestingly, $G_{Vodafone_AS}$ only shows two of the four large International Internet Service Providers (with a high Customer Cone Size) that Airtel and Bharti Airtel indicated, namely Level 3 Communications Inc. (AS3356:4160) and Cogent Communications (AS174:4541). Our evidence indicates that $G_{Vodafone_AS}$ only makes use of a few Autonomous Systems with a high Customer Cones ranking for its complete

internetworking activities.

Looking at the BGP Route Propagation graphs for the Autonomous System of Vodafone in Hurricane Electric (2016), one can see that its upstream connectivity is reliant on only one connection to Vodafone India Ltd. (AS55410:157). From here, it usually connects to Bharti Airtel Ltd. (AS9498:537), Tata Communications Ltd. (formerly VSNL), (AS4755) and Cable and Wireless Worldwide plc. (AS1273:355), where it finally reaches some of the other large Autonomous Systems mentioned in Table 4-24 below.

Vodafone highest k-core Autonomous Systems ranked by Customer Cone Size		
ASN : Transit Degree (CAIDA, 2016a)	AS- Name (Hurricane Electric, 2016)	CCS (CAIDA, 2016a)
AS3356:4160	Level 3 Communications Inc.	29,494
AS174:4541	Cogent Communications	23,299
AS2914:1187	NTT America Inc.	18,991
AS1273:355	Cable and Wireless Worldwide plc.	5,878
AS3491:540	PCCW Global	3,572
AS7018:2320	AT&T Services Inc.	3,292
AS55410:157	Vodafone India Ltd.	293
AS4538:21	China Education and Research Network Center	18
Key AS: Autonomous System. ASN: Autonomous System Number. CCS: Customer Cone Size.		

Table 4-24: Vodafone customer cones.

Furthermore, according to the CAIDA (2016b) AS relationship dataset, it is worth to notice that only one of the Vodafone AS relationships between those Autonomous Systems identified in the highest *k-core*, is of a peering nature. This relationship is the one existing between Cable and Wireless Worldwide plc. (AS1273:355) and Cogent Communications (AS174:4541), see column 4 last row in Table 4-25 below. The analysis of the merged datasets shows additional interesting relevant features for $G_{Vodafone_AS}$'s

upstream Internet market structure. We reveal new Autonomous System relationships from an actively measured Internet periphery perspective that do not appear as direct business relationships in the CAIDA (2016b) AS Relationships dataset.

Vodafone AS Relationships using CAIDA (2016a)			
SourceASN: Transit Degree, (AS Rank)	TargetASN: Transit Degree, (AS Rank)	Edge Weight	CAIDA AS Relationship (CAIDA, 2016b)
AS55410:157, (139)	AS55410:157, (139)	8,909	#N/A
AS1273:355, (12)	AS1273:355, (12)	3,360	#N/A
AS4538:21, (1151)	AS55410:157, (139)	1,772	#N/A
AS174:4541, (2)	AS174:4541, (2)	1,654	#N/A
AS55410:157, (139)	AS1273:355, (12)	1,576	#N/A
AS1299:1010, (3)	AS1299:1010, (3)	888	#N/A
AS3356:4160, (1)	AS3356:4160, (1)	420	#N/A
AS55410:157, (139)	AS3356:4160, (1)	353	#N/A
AS55410:157, (139)	#N/A	336	#N/A
AS1273:355, (12)	AS174:4541, (2)	309	p2p
Key AS: Autonomous System. ASN: Autonomous System Number. CCS: Customer Cone Size.			

Table 4-25: Vodafone AS Relationships.

Again, merging the CAIDA (2016b) AS relationship dataset with the collected Vodafone internetworking data, we can see that 1.48% of the collected AS relationships were of *customer-to-provider* nature. Additionally, 2.45% of the AS relationships represented *peer-to-peer* relationships and only 1.17% *provider-to-customer* ones. Moreover, 64.85% of all AS relationships represented AS-internal connections (*traceroute* hops from the dataset generated from the Aircel operator directly linking a source with a target IP addresses, both belonging to the same Autonomous System), and 8.73% of all AS relationships resulted from *traceroute* terminations. Finally, 21.43% of all observed

business relationships do not appear in the CAIDA (2016b) AS-Relationship dataset, see Table 4-26 below.

Vodafone AS Relationships Overview using CAIDA (2016b)				
Type of AS Relationship (CAIDA, 2016b)	Description	Number of Edge observations	Edge Weights (Traversals)	In percentage of total edge weights
-1	cp2 relationship	20	453	1.48%
0	p2p relationship	47	719	2.35%
1	p2c relationship	186	358	1.17%
#N/A	CAIDA (2016b) undetected	1493	6,564	21.43%
#N/A	AS-internal edge	676	19,865	64.85%
#N/A	<i>Traceroute</i> Termination	1197	2,674	8.73%
<p>Key</p> <p>AS: Autonomous System.</p> <p>c2p: Customer-to-provider relationship.</p> <p>p2c: Provider-to-customer relationship.</p> <p>p2p: Peer-to-peer Relationship.</p> <p>#N/A: No relationship in CAIDA (2016b) available.</p>				

Table 4-26: Vodafone AS Relationships overview.

Just like before, plotting the merged data into a two-dimensional Euclidean Space in Figure 4-23 below reveals that Vodafone is seemingly making more use of *peer-to-peer* and *customer-to-provider* relationships than both *provider-to-customer* ones. Compared to G_{Aircel_AS} and $G_{Bharti_Airtel_AS}$, the graph visualisation of $G_{Vodafone_AS}$ indicates two AS relationships with a strong edge weight, whose economic nature remains hidden. These connections are clearly established between China Education and Research Network Center (AS4538:21) and Vodafone India Ltd. (AS55410:157), and between Vodafone India Ltd. (AS55410:157) and Cable and Wireless Worldwide plc. (AS1273:355).

$G_{Vodafone_AS}$

Edge thickness: 0.05

Red edges: 1 (p2c link)

Green edges: 0 (p2p link)

Blue edges: -1 (c2p link)

Yellow edges: #N/A

(link not available)

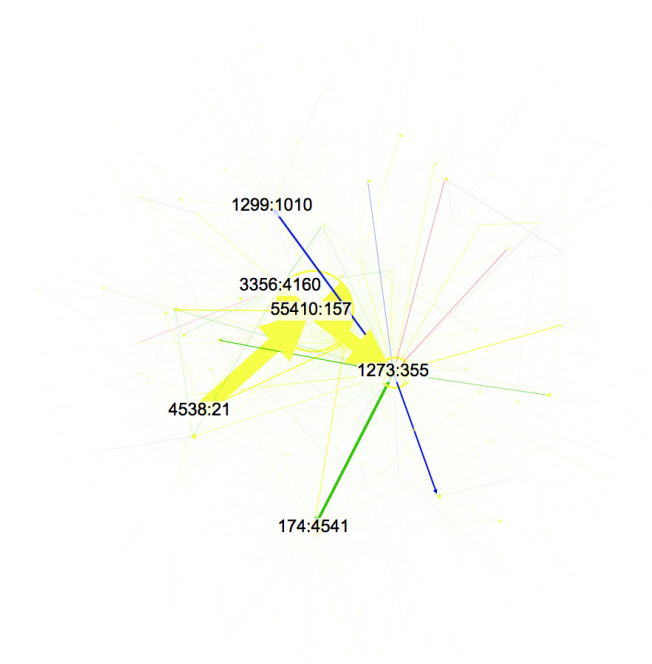


Figure 4-23: Vodafone graph visualisation with relationship colouring.

Summarising, the evidence showed that the upstream connectivity in $G_{Vodafone_AS}$ is strongly reliant on three ASes, namely Vodafone India Ltd. (AS55410:157), their Cable and Wireless Worldwide plc. (AS1273:355) subsidiary and Level 3 Communications (AS3356:4160).

Unlike the BGP Route Propagation graphs in Hurricane Electric (2016), the analysis exposed a direct connection to Level 3 Communications Inc. (AS3356:4160). This is particularly interesting since this AS represents a major Tier-1 *Internet Service Provider*. Just like before, this analysis revealed some hidden routing and structuring features of Vodafone's operator network. These features were not observable through the previously available datasets.

4.4.4 Summary Autonomous System Relationships

This section has revealed the most relevant AS vertices, the key upstream relationships as well as the hierarchical features for the three Tamil Nadu Mobile mobile broadband operators. Aircel seems to be strongly reliant on Bharti Airtel Ltd. as well as Tata Communications (formerly VSNL), both national operators. Bharti Airtel indicates a reliance on another AS of Bharti Airtel Ltd. and shows a direct connection to Level 3 Communications Inc. Finally, Vodafone shows reliance on another Vodafone India Ltd. AS but also on their International Cable and Wireless Worldwide plc. subsidiary. While this case study notes that both the IPv4 Route Propagation graphs of Hurricane Electric (2016) and the CAIDA (2016b) AS relationship data highlight gaps, it also points towards the importance of studying mobile broadband networks by utilising *traceroute* observations from active Internet periphery measurements. Moreover, we detected that Vodafone is making stronger use of *peer-to-peer* relationships than the other two sub-networks.

4.5 Summary Complex Network Analysis

The first part of this chapter covered the *Descriptive* and *Complex Network Analysis* followed by a *Graph Visualisation Analysis* at Internet Protocol granularity. Based on the observed *Clustering Coefficients* of all three mobile broadband operator networks, our analysis showed low vertex interaction intensities but *power-law degree distributions* indicating features typical of *Scale-Free Networks*. We then tested this property using graph visualisation analysis. Next, we decomposed the networks into different subsets, the *k-cores*, based on the *k-core decomposition*, which revealed the most important IP address vertices that the operator networks relied upon for upstream internetworking connectivity towards the network periphery.

The second part of this chapter covered the *Complex Network*, the *Graph Visualisation analysis* and Centrality metrics at Autonomous System granularity. Here, again we observed the *Clustering Coefficient* for the three operator networks, the previous finding that the three operator networks follow *Scale-Free Network* models. We also explored the *Eigenvector Centrality*, of the relevant networks, finding that it represents a key additional Centrality metric for the analysis of *traceroute* measurements. Just like at IP granularity, the *k-core decomposition* algorithm revealed the most important Autonomous Systems that the three Tamil Nadu mobile broadband operator networks relied upon for their respective upstream internetworking connectivity.

In the last part of this chapter, these Autonomous Systems were further analysed by merging our primary data with secondary CAIDA datasets detailing their inferred bilateral economic relationships.

4.6 Key Findings Complex Network Analysis

This chapter provides some significant findings, highlighting parts of our contribution to knowledge. First, we showed the value of analysing upstream active Internet periphery measurements data from two levels of granularity. In particular, the analysis at AS granularity showed interesting hidden features for the three Tamil Nadu mobile broadband operator networks. Next, this dissertation provides a pioneering case study linking primary upstream connectivity data using active Internet periphery measurements with secondary CAIDA (2016b) AS-relationship data. This connection of two datasets revealed some previously undiscovered AS relationships. Furthermore, we indicated that the three mobile broadband operators are reliant upon a number of densely connected upstream Autonomous Systems of large, mostly Tier-1, *Internet Service Providers*. Our exploratory approach to research, using a *Descriptive Analysis*, *Complex Network Analysis*, *Graph Visualisation Analysis*, and dataset fusion and filtering provided a fruitful combination of research methods to understand the key role played by the different agents composing the upstream Internet infrastructure, and to improve our understanding of the complex nature of the Tamil Nadu mobile broadband operator networks.

5 STATISTICAL NETWORK ANALYSIS

This chapter covers the *Statistical Network Analysis* of the three Tamil Nadu mobile broadband operator networks discussed from a *Descriptive* and *Graph Visualisation Analysis* perspective in the previous Chapter 4. This chapter will introduce four different econometric models to test the elaborated *Working Hypotheses*, abducted through the Literature Review in section 2.5. Hence, we commence this chapter with a review of our *Working Hypotheses*, alongside a general overview of our two-stage econometric estimation process used to study the relationship between upstream hierarchical upstream Internet market structuring and mobile broadband affordability (measured in price per Megabyte).

We argue that this relationship is of economic importance, e.g. when assessing the supply-side of mobile broadband connectivity and the resulting affordability of infrastructural access from an end-user perspective (smartphone users). Stronger hierarchical upstream Internet market structures indicate that a few large *Internet Service Providers* act as connectivity bottlenecks with strong downstream price and bargaining powers. Hence, we anticipate that a stronger hierarchical upstream Internet market structure of a mobile broadband operator would lead to higher price per Megabytes of price plans, an indicator for mobile broadband affordability for (especially data sensitive) end-users. This is of particular importance for end-users in more rural areas of Tamil Nadu, given the large present income inequalities identified by TN-GOV-IN (2014), Sundar (2015) and Selvabaskar et al. (2016).

In the first stage of our two-stage estimation process, we specify two econometric models (Model 1 and Model 2) to estimate the relationship between the network structuring markers (*Clustering Coefficient* and *Eigenvector Centrality*) and the upstream connectivity (*Weighted Out- and In-Degree*) for each one of the three Tamil Nadu mobile broadband operator networks. We will then use these associations' estimated coefficients as proxies, representing each mobile broadband operator networks' levels of hierarchical structuring.

In the second stage, we use these proxies (estimated in the first stage), as an additional explanatory variable, to estimate the effects of each mobile broadband operator networks'

hierarchical structuring on our key dependent variable of interest – the price per Megabyte. This metric shall represent end-user affordability of each mobile broadband operators’ advertised price plans. These price plans were derived from a secondary dataset (see section 5.4). This chapter concludes by presenting a correlation between the obtained levels of hierarchical structuring for the three Tamil Nadu mobile broadband operator networks and Quality of Service (QoS) metrics obtained from the Telecom Regulatory Authority of India.

5.1 Review of Working Hypotheses

The *Working Hypotheses*, abducted in the Literature Review (see section 2.5), provide a useful guideline for this chapter. Here, the identification of hierarchical structuring for WH1, WH1.1 and WH1.2 will be covered in Model 1 and Model 2, the first stage of our two-stage econometric estimation process. WH2 is then analysed in Model 3 and Model 4 in the second stage. WH3 itself is solely analysed using a correlation analysis in the second stage. Figure 5-1 below provides an overview of all econometric models employed in this chapter.

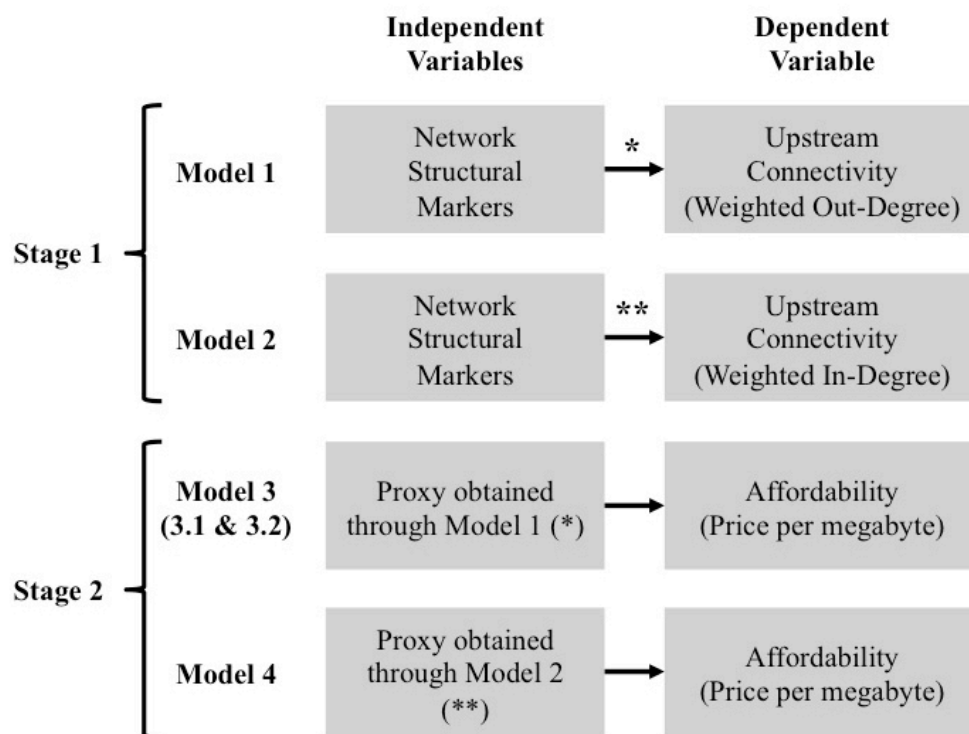


Figure 5-1: Overview of econometric models.

5.2 Background

The theoretical background for our econometric models is given by the choice of Complex Network metrics discussed in Chapter 4. Complex Networks are non-trivial systems that often display *power-law degree distributions*. As such, only certain vertex metrics may help in explaining the Internet market access structure for the three Tamil Nadu mobile broadband operators' connectivity networks of this case study.

The connectivity of a network graph is, according to Kolaczyk (2009), associated with the flow of information within it. Every vertex in a directed graph plays two roles. First, the broadcaster and second, the receiver of information between a set of interconnected Autonomous Systems (Benzi, 2014, p.62). These roles, and hence a measurement for connectivity, are provided by the *Out-Degree* and *In-Degree* values of a vertex, which suits well to the nature of our directed graphs. Measuring the structure of connectivity in a network may be covered by many different centrality measures that correspond to different notions of interest. However, and given their theoretical grounding, most of these metrics, alone, are unable to capture the properties of interest in this case study. *Degree Centrality* e.g. does not take vertex importance as well as edge directions (as employed by *traceroutes*) into consideration. The often-used *Closeness* and *Betweenness Centralities* assume that communication in a network always follows the shortest paths, which is unlikely to be the case for the upstream Internet market. Here, the connectivity of *traceroutes* would not necessarily follow the shortest connectivity paths.

The spectral measure of the Bonacich (1987) *Eigenvector Centrality* will be the key centrality metric used in the following analysis of the mobile broadband operator networks. This metric takes the influence of a vertex in a given network graph into consideration (Newman, 2006). Additionally, the *Eigenvector Centrality* is sensitive to situations where vertices with a low degree are connected to those with higher degrees, or vice versa (Bonacich, 2007, p.561). This seems apparent in our networks, as previously indicated in Chapter 4. Moreover, the *Eigenvector Centrality* considers both, direct connections between any pair of vertices and indirect connections to other network vertices. According to Bonacich (2007, p. 564), the *Eigenvector Centrality* is distinctively appropriate where vertex centrality is given by *Degree* differences. Utilising this metric for studying upstream connectivity networks may, therefore, reveal a positive effect of the *Eigenvector Centrality* on connectivity. This inference is backed by the specific network characteristics of the three mobile broadband operators discussed in the

Descriptive and Graph Visualisation Analysis in Chapter 4. We assume that certain Autonomous Systems have a strong influence on other Autonomous Systems (captured by the *Eigenvector Centrality*), resulting in a more hierarchical level of the mobile broadband operator networks. Moreover, Borgatti (2005, p.56) notes that the *Eigenvector Centrality* is appropriate for studying information flows in networks.

The *Clustering Coefficient* provides another network metric, essential to understand relevant features of connectivity in a given network. This metric is used to study Internet structures (Barrat et al., 2004). Moreover, scholars of Internet structuring argue that the hierarchical organisation of a network may be captured by estimating the parameters that represent the effect of the *Clustering Coefficient* on the *Degree* connectivity in a given network (Vázquez, Pastor-Satorras, and Vespignani, 2002). These authors also found that the *Clustering Coefficient* scales, with a negative exponent, as a *power-law* function of the networks' level of connectivity. This indicates that those Autonomous Systems that connect to a few, but larger, Autonomous Systems, are not well connected amongst each other. This, in itself, indicates a hierarchical structuring and therefore potential structural bottlenecks of the upstream Internet access layers. Moreover, as far as we are aware, there is a gap in the literature in using the *Eigenvector Centrality* alone, or in combination with the *Clustering Coefficient*, to explain hierarchical structuring of upstream connectivity in mobile broadband networks (see section 2.3.3).

5.3 Hierarchical Structuring

Given the structural indicators discussed in Chapter 4, this section looks at the possible effects of the hierarchical structuring markers on connectivity. In detail, Model 1 first analyses the effect of these hierarchical structuring markers on the *Weighted Out-Degree* connectivity, whereas Model 2 similarly analyses the effect of the hierarchical structuring markers (*Clustering Coefficient* and *Eigenvector Centrality*) on the *Weighted In-Degree* connectivity (see section 5.3.2).

The combination of these relationships studied in Model 1 and Model 2 are used to explore the following *Working Hypotheses* derived from the Literature Review (see section 5.1 above):

WH1: ‘The Tamil Nadu mobile broadband operators’ upstream Internet market structure displays features of a hierarchical ordering’.

WH1.1: ‘The Tamil Nadu mobile broadband operators rely on an identified set of specific Internet Service Providers for their upstream connectivity’.

5.3.1 Descriptive Statistics

The Autonomous System observations, from the primary active Internet periphery measurements using the Portolan (2015) Android application, were chosen as the unit of analysis. This choice was grounded in the belief that vertex observations at Autonomous System granularity, compared to edge observations, were more useful in capturing those structural properties of interest in this case study. Intrinsically, however, these vertex observations embody metrics being calculated using the directed edge metrics in Chapter 4. The dataset, therefore, covered 2,215 unique vertex observations. While 68.30% of these observations captured Autonomous Systems being collected through the Vodafone SIM card, only 23.57% of the observations were collected from the Aircel SIM card, and only 8.13% from the Bharti Airtel one, respectively. Considering this fragmentation, the following analysis, when covering the total dataset, would have revealed results being highly influenced by Vodafone. We, therefore, focused on per-operator separated datasets to obtain the necessary insights. Moreover, the relevant variables were chosen, by considering the emerging insights discussed on various occasions in the Literature Review.

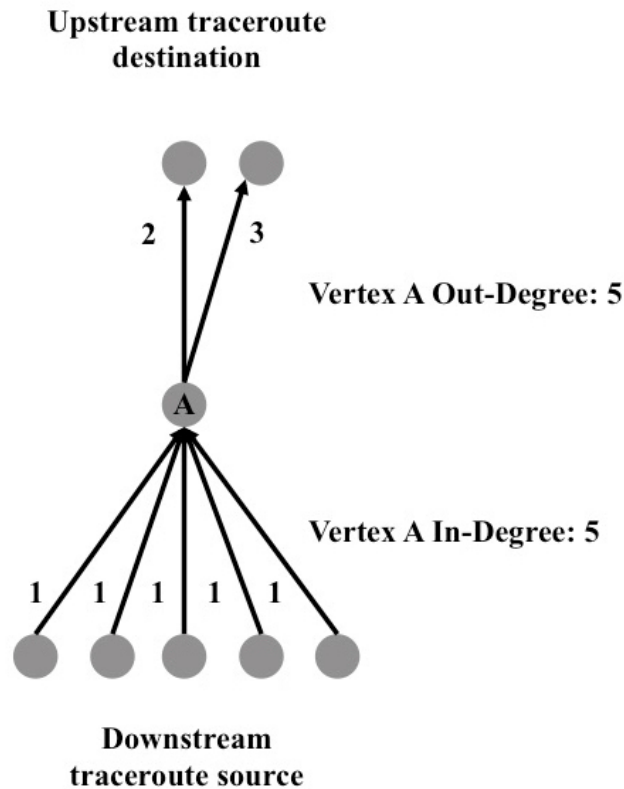
In the following, the operator vertex observations from Chapter 4 are filtered per operator so that the following econometric models will be performed separately, for each of the mobile broadband operators, based on the primary data collected from their specific SIM cards. This allows to later compare the operator-based differences emerging from the econometric models. After importing the initial primary dataset into Stata (2016), Table 5-1 below reports the ranges for the variables that will be used (after suitable logarithmic transformation) in the econometric models below.

Descriptive statistics					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
woutd	2215	16.40542	275.2919	0	11,394
winde	2215	16.40542	264.3536	0	10,686
clus	2215	.2827968	.1998676	0	.6666667
eige	2215	.0027941	.0401714	0	1
Key clus = Clustering Coefficient. eige = Eigenvector Centrality. woutd = Weighted Out-Degree. winde = Weighted In-Degree.					

Table 5-1: Descriptive statistics.

An analysis using the Stata (2016) Data Browser revealed that the observations covered a few Autonomous Systems with relatively large *Weighted In-Degrees* (*winde*) and *Weighted Out-Degrees* (*woutd*), while most of the Autonomous Systems showed lower values, representing the nature of *power-law degree distributions* (see Figure 4-12).

Given the features of *traceroutes* generated from the Internet periphery (smartphones used in the data collection (see section 3.3.4)), an Autonomous System located along this *traceroute's* path, is likely to have a large *Weighted Out-Degree* (*woutd*). This represents a large set of different outgoing next hop destinations when it has a large *Weighted In-Degree* (*winde*), when it is reached from a large set of different incoming connections originating potentially from many different Autonomous Systems. Figure 5-2 below illustrates this concept.



Key

- Vertex without label.
- Directed edge, linking a pair of vertices.

Figure 5-2: In-Degree and Out-Degree flow through a vertex.

Table 5-2 below provides the key *Descriptive Statistics* for the variables of interest, *Weighted Out-Degree*, *Weighted In-Degree*, *Clustering Coefficient* and *Eigenvector Centrality*, for the three mobile broadband operators (1 = Aircel, 2 = Bharti Airtel, 3 = Vodafone).

Descriptive statistics by mobile broadband operator						
Operator	Variable	Obs.	Mean	Std. Dev.	Min.	Max.
1	Woutd	522	9.097701	83.35425	0	1847
1	Winde	522	9.097701	75.67523	0	1581
1	Clus	522	.2375005	.1961916	0	.5833333
1	Eige	522	.0026668	.0438957	0	1
2	Woutd	180	5.311111	18.40916	0	195
2	Winde	180	5.311111	17.31802	0	148
2	Clus	180	.2242081	.1988603	0	.5
2	Eige	180	.0184107	.0893504	0	1
3	Woutd	1513	20.24653	329.3778	0	11394
3	Winde	1513	20.24653	316.6572	1	10686
3	Clus	1513	.3053947	.1972844	0	.6666667
3	Eige	1513	.0009801	.0268621	1.12e-06	1
Key clus = Clustering Coefficient. eige = Eigenvector Centrality. operator (mobile broadband operator): 1 = Aircel, 2 = Bharti Airtel, 3 = Vodafone. woutd = Weighted Out-Degree. winde = Weighted In-Degree.						

Table 5-2: Descriptive statistics by mobile broadband operator.

Looking at those Autonomous Systems that are showing an especially large or low *Weighted Out-Degree* (*woutd*) or *Weighted In-Degree* (*winde*) revealed an interesting pattern regarding outlier observations in the dataset (covering all three operators). According to the analysis of Table 5-3 on the next page, the mobile broadband operator networks indicate only two data points with a large *Weighted Out-Degree* and *Weighted In-Degree* simultaneously.

Weighted In-Degree and Weighted Out-Degree Matrix		
	Autonomous System with a small Weighted In- Degree, (operator)	Autonomous System with a large Weighted In- Degree, (operator)
Autonomous System with a small Weighted Out- Degree, (operator)	Most observations.	N/A (1)* N/A (2)* N/A (3)*
Autonomous System with a large Weighted Out- Degree, (operator)	AS4538 (3) AS4755 (1)	AS55410 (3) AS1273 (3) AS174 (3) AS6453 (1) AS1299 (3) AS3356 (3) AS4755 (1) AS4538 (3) AS45609 (2)
<p>Key</p> <p>* normal terminations or drops of a given <i>traceroute</i> measurement.</p> <p>operator (mobile broadband operator) 1 = Aircel, 2 = Bharti Airtel, 3 = Vodafone.</p> <p>Note: Inside the parenthesis, we report the number referring to the mobile broadband operator, from which SIM-card we generated the respective measurements.</p>		

Table 5-3: Highest and lowest Weighted In-Degree and Weighted Out-Degree matrix.

This is furthermore visualised in the following Figure 5-3, where plotting the *Weighted In-Degree* (*winde*) against the *Weighted Out-Degree* (*woutd*) for all mobile broadband operators revealed the identified outliers, and the overall high correlation coefficient valued at 0.9636. Looking at the Aircel outliers in Figure 5-4 below indicated one outlier with a large *woutd* and a low *winde*. A manual analysis using the Stata (2016) Data Browser identified this vertex as Tata Communications (formerly VSNL), (AS4755), one of the outliers indicated in Figure 5-3 below. The correlation coefficient was set at 0.9184. Plotting the two-way scatter plot for the *winde* against *woutd* for the Bharti Airtel observations in Figure 5-5 below revealed none of the outlier observations from Figure 5-3 below, while the correlation coefficient was lower with 0.6542. Nevertheless, Figure 5-5 shows a number of outliers for the Bharti Airtel observations that only emerge in this figure, given their relative smaller deviations from the norm than those observed for all observations as depicted in Figure 5-3.

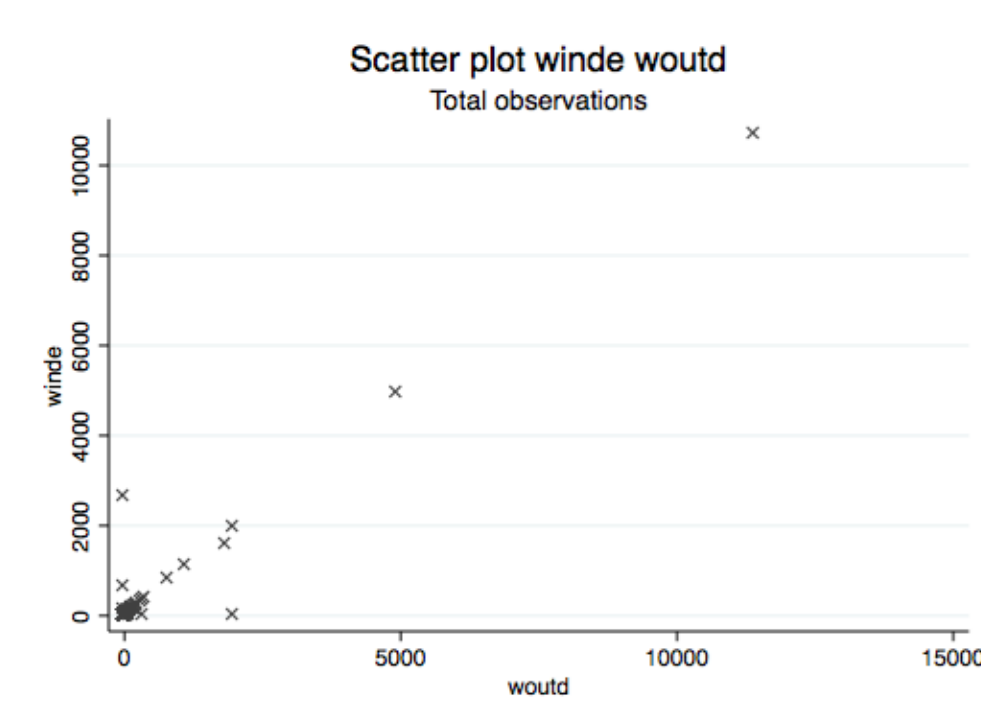


Figure 5-3: Total observations scatter plot *Weighted In-Degree* against *Weighted Out-Degree*.

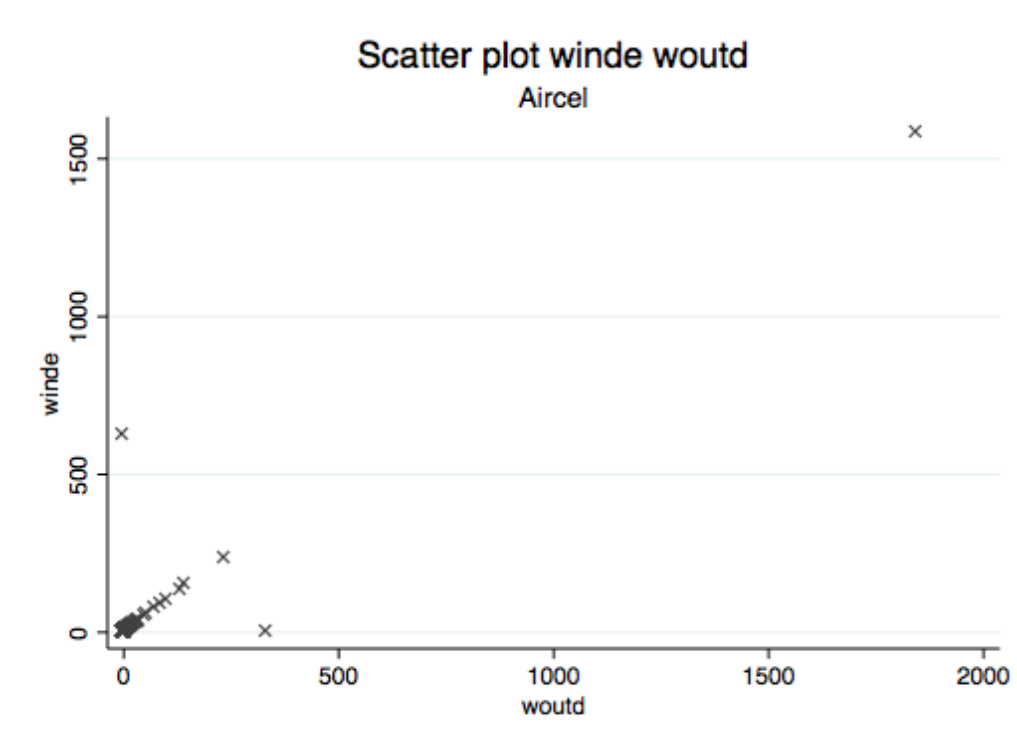


Figure 5-4: Aircel scatter plot Weighted In-Degree against Weighted Out-Degree.

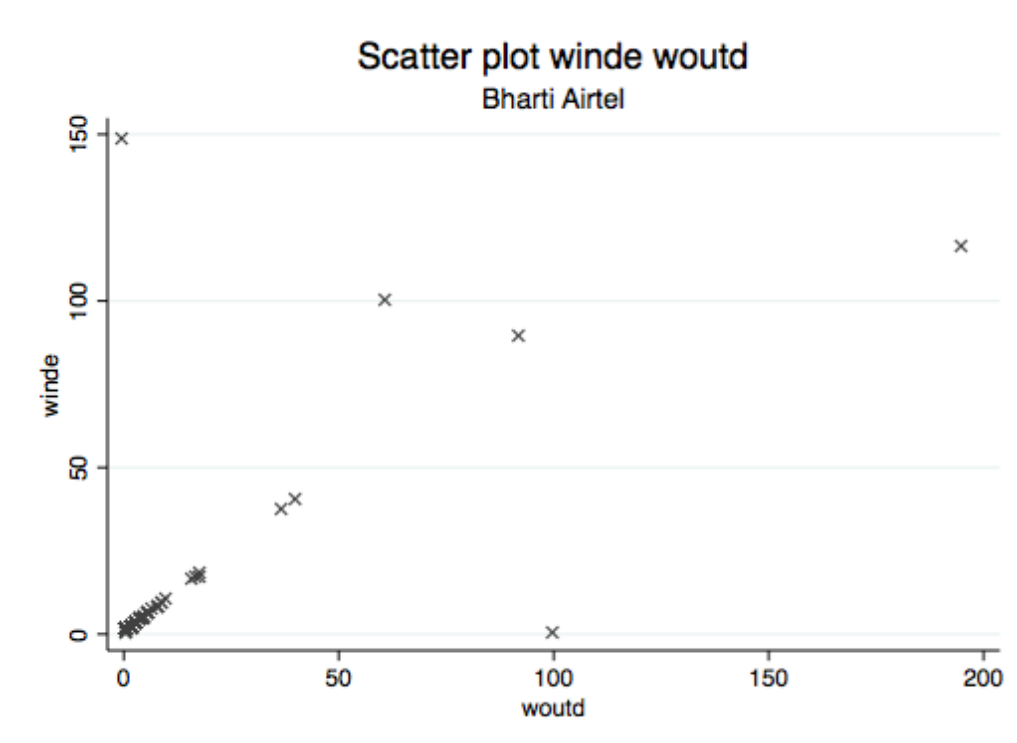


Figure 5-5: Bharti Airtel scatter plot Weighted In-Degree against Weighted Out-Degree.

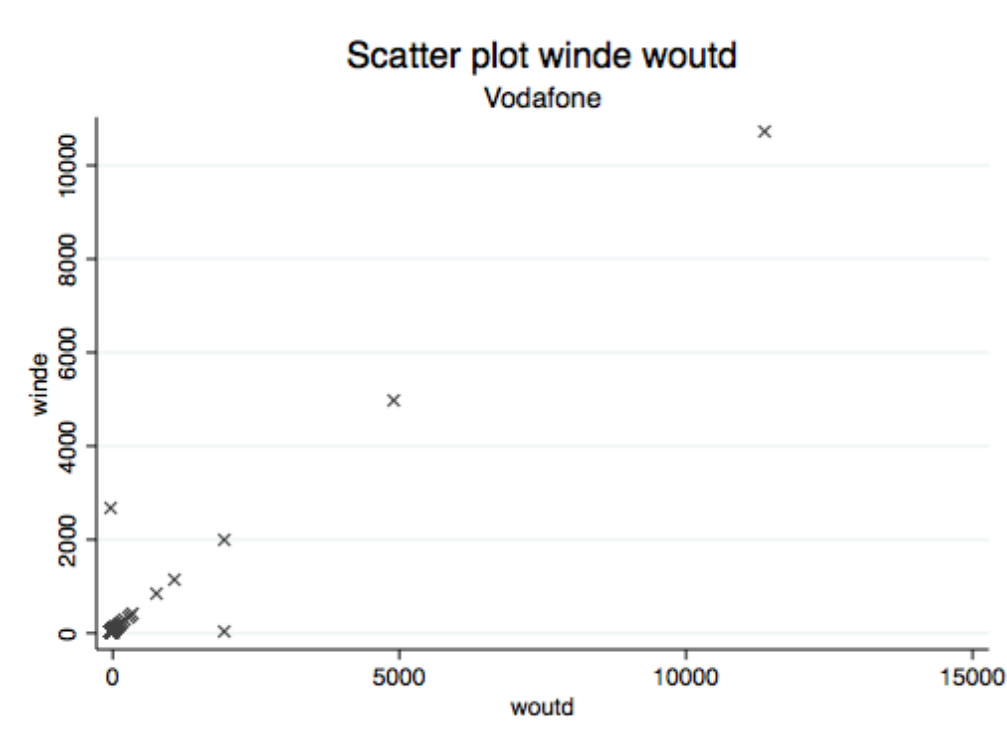


Figure 5-6: Vodafone scatter plot *Weighted In-Degree* against *Weighted Out-Degree*.

Finally, in plotting *winde* against *woutd* for the Vodafone observations in Figure 5-6 above indicated another outlier from Figure 5-3 (China Education and Research Network (AS4538)) and a high correlation coefficient of 0.9647.

Interestingly, both of the identified vertices, Tata Communications (formerly VSNL), (AS4755) and the China Education and Research Network (AS4538) were already identified as structurally important Autonomous Systems in section 4.3 above. This indicates that these two Autonomous Systems are potentially used by the respective mobile broadband operators for linking to other Autonomous Systems, probably to those ones for which they lack a direct connection themselves. Looking more closely at those Autonomous Systems that incorporated a high *woutd* and *winde* in Table 5-3 above, revealed those Autonomous Systems that apparently featured as the core of the network. Hence, these Autonomous Systems are potentially the first hop of a given *traceroute* observation for each of the three operators. A visual analysis with the Stata (2016) Data Browser showed that the most notable Autonomous System for Aircel observations was Tata Communications (America) Inc. (AS6453). Bharti Airtel only showed one strong Autonomous System being their own Bharti Airtel Ltd. (AS45609), while Vodafone showed several ones, namely Vodafone India Ltd. (AS55410), Cable and Wireless Worldwide plc. (AS1273), Cogent Communications (AS174), Telia Company AB

(AS1299) and Level 3 Communications (AS3356). These descriptive results came as no surprise, given the previous findings in Chapter 4.

Therefore, those Autonomous Systems indicating high *woutd* and high *winde* seem to be primarily responsible for parts of the network structural organisation, and hence, amongst others, structural bottlenecks, for the network originating from each mobile broadband operator SIM card. Similarly, those Autonomous Systems with a high *Weighted Out-Degree* (*woutd*) and a low *Weighted In-Degree* (*winde*) are potentially additional drivers for structural bottlenecks. Based on these findings, the following econometric models in this chapter should reveal diverging levels of hierarchical structuring for the three mobile broadband operators. Given the identified outlier Autonomous Systems, hierarchical organisations and hence structural bottlenecks should be more apparent for Aircel and Vodafone than for Bharti Airtel.

Looking, in more detail, at the Descriptive Statistics for the covariates revealed additional insights (see Table 5-1). Plotting the histogram of the *clus* distribution in Figure 5-7 below indicated that a large number of observations incorporated either no *clus*, or fairly large values.

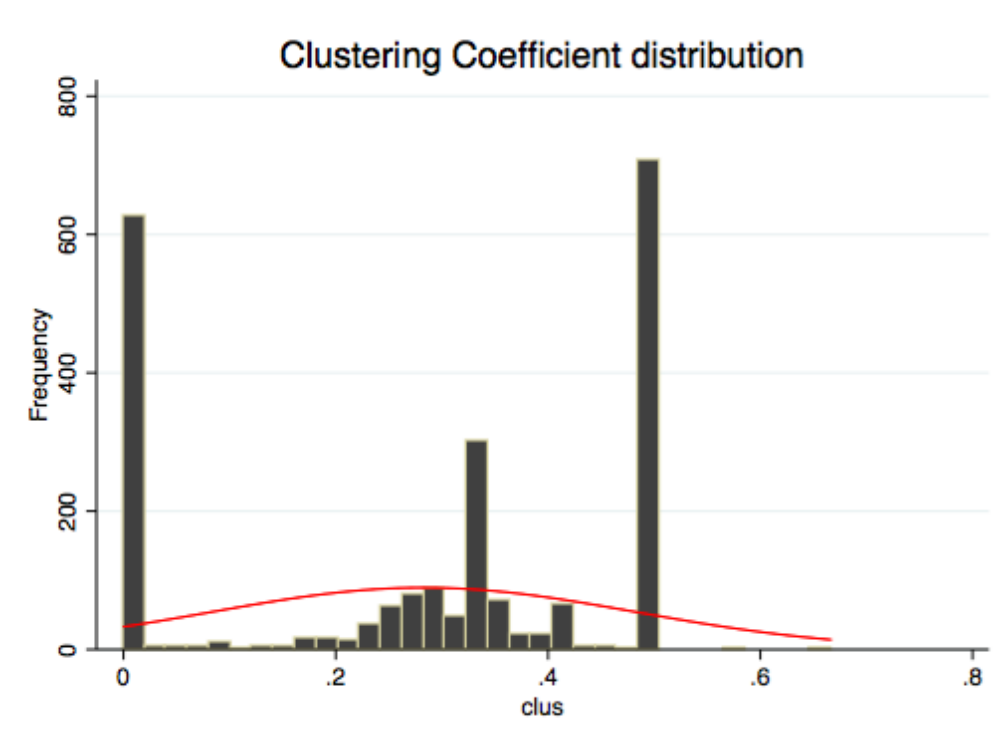


Figure 5-7: Clustering Coefficient distribution with fit line.

Looking at the Autonomous Systems with the largest *clus* values in the following Table 5-4 revealed additional insights. The data originated from Aircel showed BIGLOBE Inc. (AS55394) having the highest *Clustering Coefficient* (.5834), followed by a multitude of other ASes with a *Clustering Coefficient* of .5.

Bharti Airtel showed also a multitude of Autonomous Systems with a *Clustering Coefficient* of .5, including e.g. General Telecommunication Organization (AS8529).

The data originated from Vodafone showed the United States Federal Reserve Board (AS10754) as the Autonomous System with the largest *Clustering Coefficient* (.6667), followed by a number of other Autonomous Systems with a *Clustering Coefficient* of .5, including e.g. West Call LLC (AS25408).

Table of highest Clustering Coefficients (clus)		
Aircel Autonomous System observation, (clus)	Bharti Airtel Autonomous System, (clus)	Vodafone Autonomous System, (clus)
AS55394, (.5834)	AS8529, (.5)	AS10754, (.6667)
AS29470, (.5)	AS11096, (.5)	AS25408, (.5)
AS38393, (.5)	AS6128, (.5)	AS20446, (.5)
AS16300, (.5)	AS12831, (.5)	AS197451, (.5)
AS8271, (.5)	AS6210, (.5)	AS48237, (.5)
AS15682, (.5)	AS6181, (.5)	AS33871, (.5)
.	.	.
.	.	.
Key clus = Clustering Coefficient.		

Table 5-4: Highest Clustering Coefficient observations.

Interestingly, those Autonomous Systems with the highest *Eigenvector Centrality* (*eige*) in Table 5-5 below, and hence with a strong network influence were, amongst others, Tata Communications (America) Inc. (AS6453), Level 3 Communications (AS3356) and Cogent Communications (AS174) for Aircel, Cogent Communications (AS174), Bharti Airtel Ltd. (AS9498) and Level 3 Communications (AS3356) for Bharti Airtel and finally Vodafone India Ltd. (AS55410), Cable and Wireless Worldwide plc. (AS1273) and Level 3 Communications for Vodafone (AS3356).

Table of highest Eigenvector Centralities (eige)		
Aircel Autonomous System observation, (eige)	Bharti Airtel Autonomous System, (eige)	Vodafone Autonomous System, (eige)
AS6453 (1)	AS174, (1)	AS55410 (1)
AS3356, (.05886702)	AS9498, (.44281486)	AS1273 (.29650135)
AS174, (.03417846)	AS3356, (.41974529)	#N/A (.04423916)
#N/A, (.02765014)	N/A, (.23068346)	AS3356 (.04212998)
A2914, (.01933738)	AS2914, (.14945413)	AS174 (.01871963)
AS12389, (.01657356)	AS34661, (.12144484)	AS3491 (.01577914)
.	.	.
.	.	.
Key eige = Eigenvector Centrality.		

Table 5-5: Highest Eigenvector Centrality observations.

Table 5-6 below displays additional *Descriptive Statistics*. The values revealed a higher Skewness for *woutd*, *winde* and *eige*, which suggested a non-normal distribution and therefore an indication that the variables ought to be transformed.

Detailed descriptive statistics			
Variable	Variance	Skewness	Kurtosis
woutd	75,785.63	34.98263	1,368.114
winde	69,882.84	33.40939	1,260.274
clus	.0399471	-.3849493	1.614403
eige	.0016137	22.13861	528.4257
Key clus = Clustering Coefficient. eige = Eigenvector Centrality. woutd = Weighted Out-Degree. winde = Weighted In-Degree.			

Table 5-6: Detailed descriptive statistics.

The difference in Kurtosis revealed the heavy tail effect as already shown in the plots of the variable distributions. Since a normal distribution would display a Kurtosis of 3, the data indicated again that only *clus* showed a somewhat normal distribution. The distributions for the other variables were of non-normal nature. Especially the *eige* distribution might follow a *power-law* function. To address the problem of non-normal distributions, the variables were transformed to their natural logarithmic base, leading to distributions of the transformed variables being closer to normal distributions (Table 5-7 below).

Ln-transformed descriptive statistics					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
lwoutd	2212	.7412221	1.080887	0	9.340842
lwinde	2207	.742654	1.082643	0	9.27669
lclus	1589	-.9811632	.3521673	-3.516868	-.4054651
leige	2207	-11.49549	2.476857	-13.70155	0
Key lclus = Ln(Clustering Coefficient). leige = Ln(Eigenvector Centrality). lwoutd = Ln(Weighted Out-Degree). lwinde = Ln(Weighted In-Degree).					

Table 5-7: Ln-transformed descriptive statistics.

Table 5-8 captures the detailed *Descriptive Statistics* for the ln-transformed variables, which appeared to be closer to a normal distribution, having a lower level of Skewness compared to before.

Ln-transformed detailed descriptive statistics			
Variable	Variance	Skewness	Kurtosis
lwoutd	1.168317	2.379546	11.38487
lwinde	1.172116	2.39158	11.4741
lclus	.1240218	-2.200099	12.23697
leige	6.13482	1.599541	5.326937
Key lclus = Ln(Clustering Coefficient). leige = Ln(Eigenvector Centrality). lwoutd = Ln(Weighted Out-Degree). lwinde = Ln(Weighted In-Degree).			

Table 5-8: Ln-transformed detailed descriptive statistics.

The following Figure 5-8 - Figure 5-11 plot the distributions for *lwoutd*, *lwinde*, *lclus* and *leige*, respectively. The distributions still indicated some left-sided Skewness for *lwoutd* and *lwinde*, while *lclus* indicated a slightly right-sided, negative Skewness.

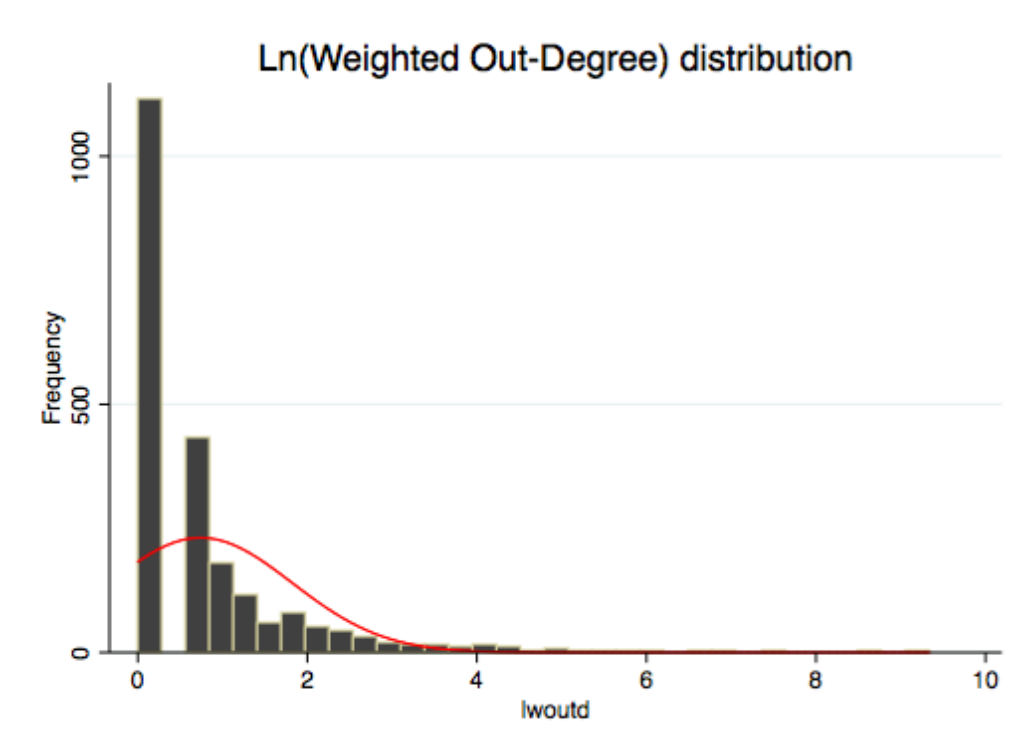


Figure 5-8: Ln(Weighted Out-Degree) distribution with fit line.

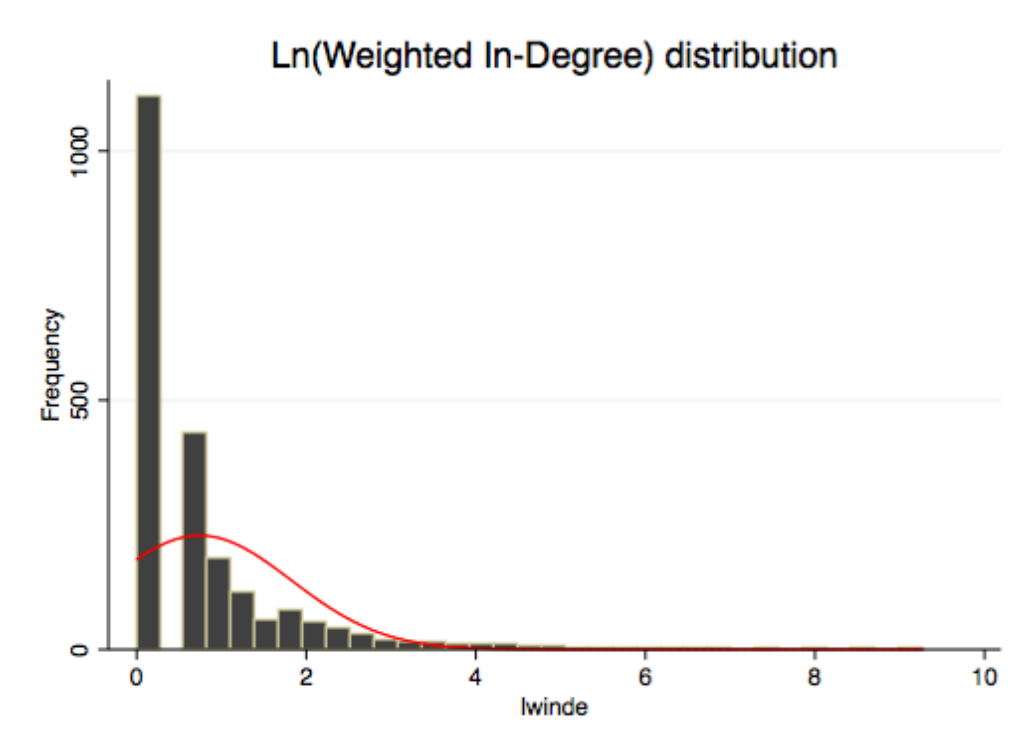


Figure 5-9: $\text{Ln}(\text{Weighted In-Degree})$ distribution with fit line.

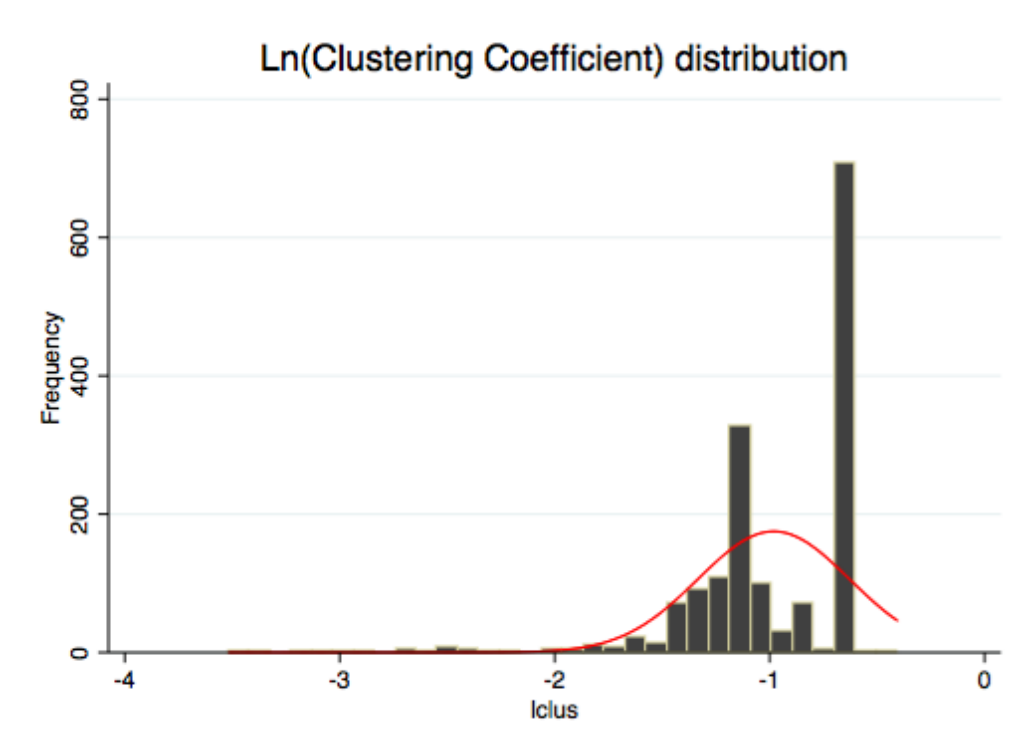


Figure 5-10: $\text{Ln}(\text{Clustering Coefficient})$ distribution with fit line.

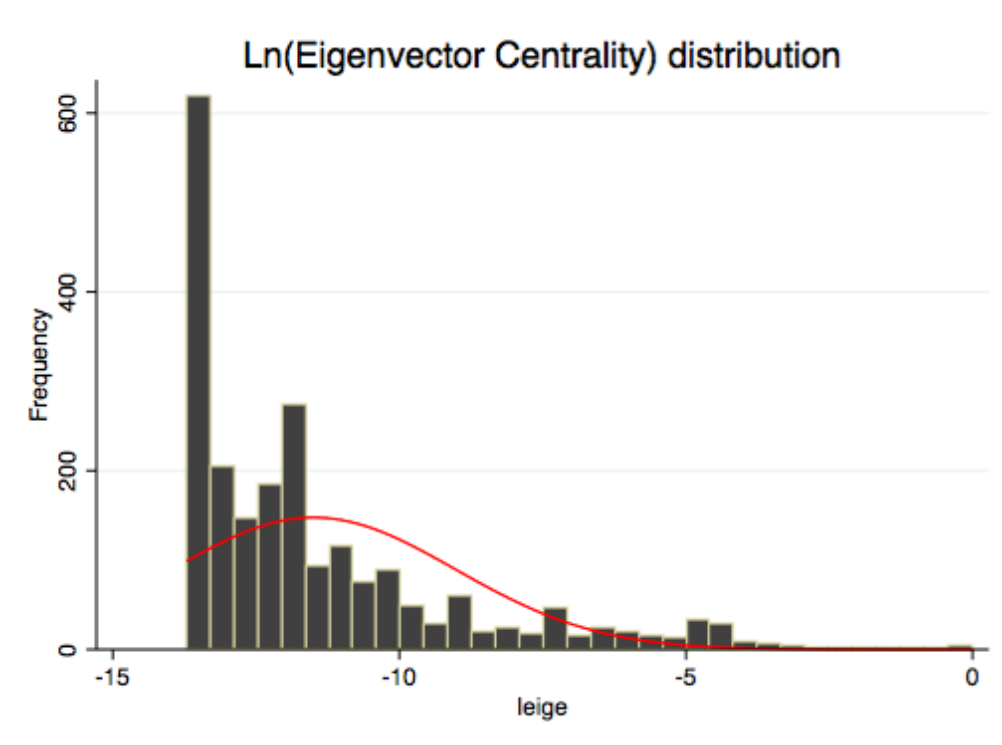


Figure 5-11: *Ln(Eigenvector Centrality) distribution with fit line.*

The *Descriptive Statistics* for the Ln-transformed variables suggest that the functional form specification might be best modelled using a log-log model. Table 5-9 below captures the correlation coefficients of the Ln-transformed variables. Interestingly, *lclus* is negatively correlated with *lwoutd*.

Ln-transformed variable correlation coefficients		
-	lclus	leige
lwoutd	-0.6271	0.5475
lwinde	0.6363	0.5629
Key lclus: Ln(Clustering Coefficient). leige: Ln(Eigenvector Centrality). lwinde: Ln(Weighted In-Degree). lwoutd: Ln(Weighted Out-Degree).		

Table 5-9: *Ln-transformed variable correlation coefficients.*

Given the great differences in observations per operator, the following plots scrutinise the data, to be later used in our econometric models, after these have been filtered according to the mobile broadband network operator of origin. In detail, we start by plotting the two variables *lclus* and *leige*, separately, by operator against *lwoutd* and against *lwinde*.

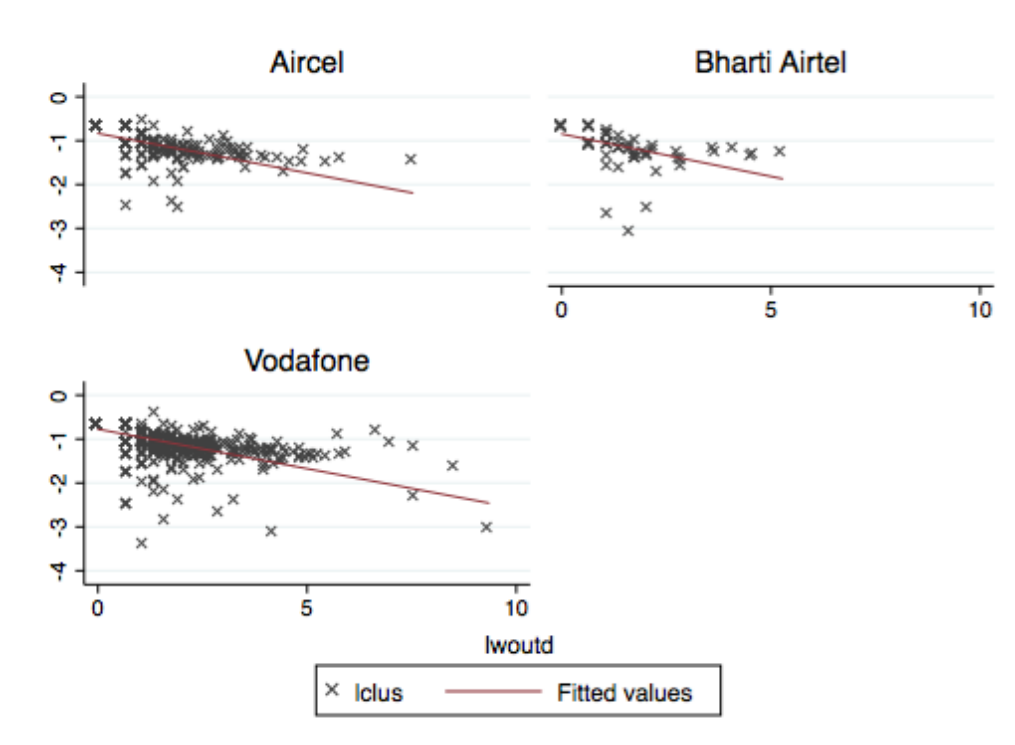


Figure 5-12: Two-way scatter plots $\text{Ln}(\text{Clustering Coefficient})$ against $\text{Ln}(\text{Weighted Out-Degree})$ per operator with linear fit line.

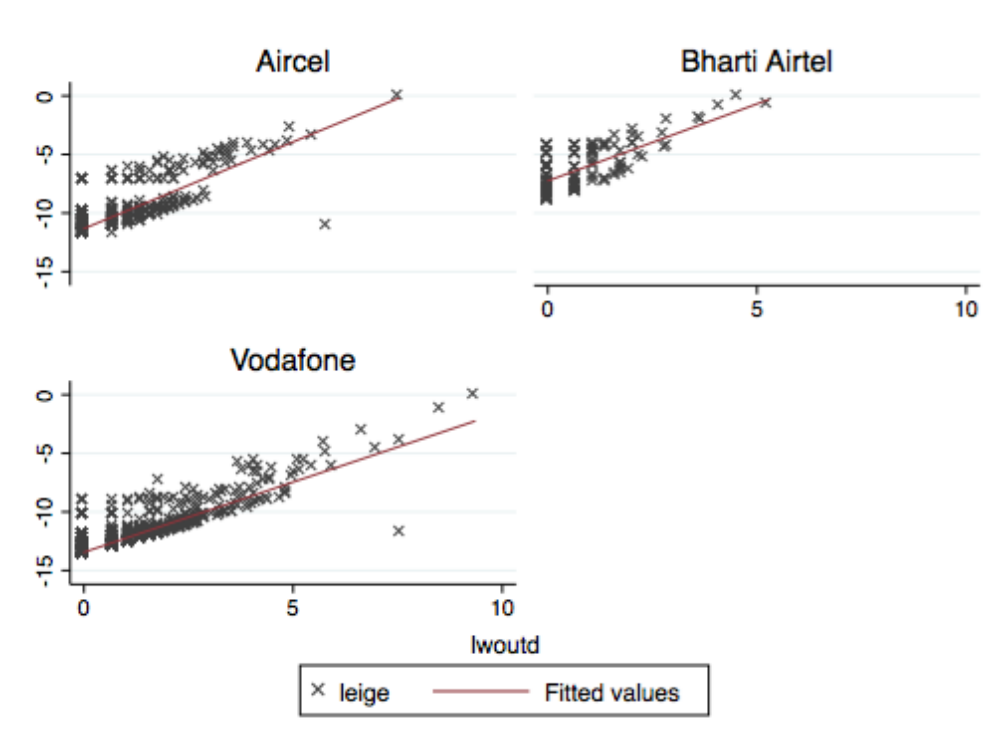


Figure 5-13: Two-way scatter plots $\ln(\text{Eigenvector Centrality})$ against $\ln(\text{Weighted Out-Degree})$ per operator with linear fit line.

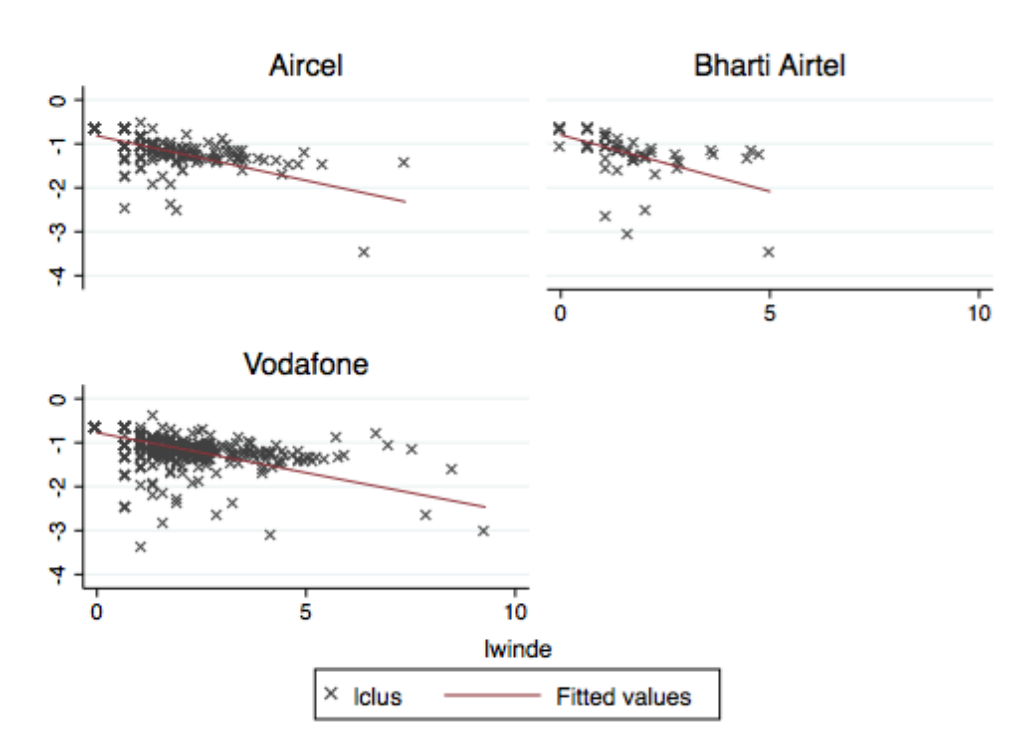


Figure 5-14: Two-way scatter plots $\ln(\text{Clustering Coefficient})$ against $\ln(\text{Weighted In-Degree})$ per operator with linear fit line.

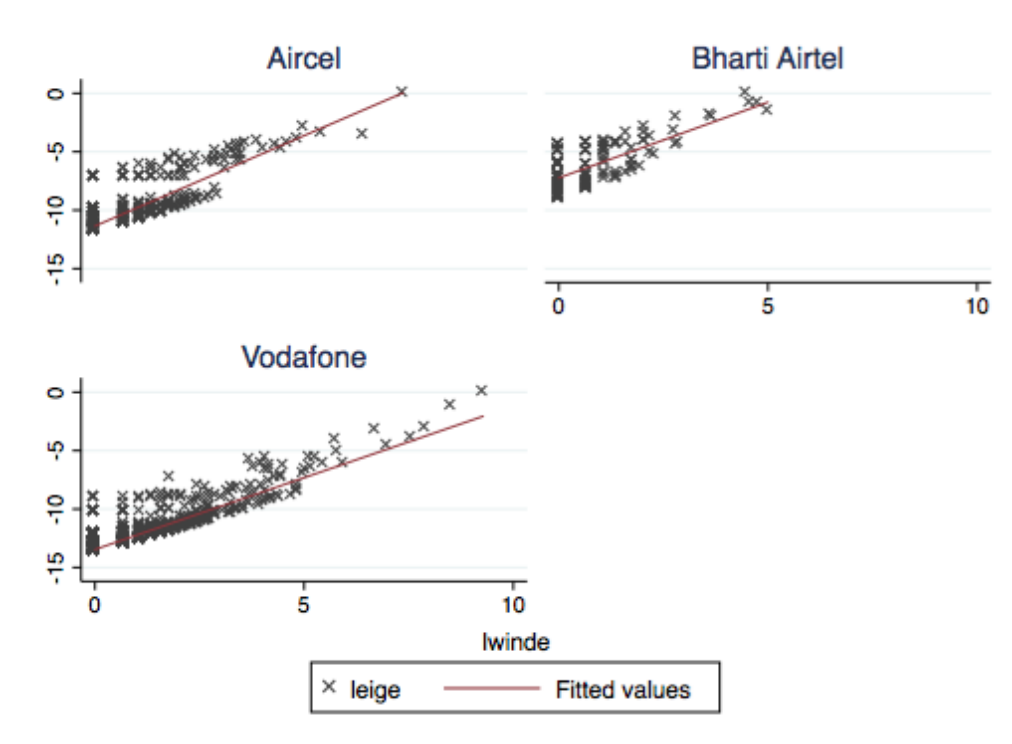


Figure 5-15: Two-way scatter plots $\text{Ln}(\text{Eigenvector Centrality})$ against $\text{Ln}(\text{Weighted In-Degree})$ per operator with linear fit line.

Based on the above plots, it seems that Vodafone shows considerably more outlier values. These could represent ASes being more responsible for hierarchical ordering than in the networks for Aircel and especially Bharti Airtel. Moreover, Vodafone also showed larger $lwoutd$ or $lwinde$ values. The same seemed to apply for the plot of the $leige$ values against the $lwoutd$ and $lwinde$ ones. Here, the $leige$ values of Vodafone were covered by much lower values than for both Aircel and Bharti Airtel.

Table 5-10 - Table 5-12 on the following pages capture the *Descriptive Statistics* for the \ln -transformed variables per mobile broadband operator (Aircel, Bharti Airtel and Vodafone, respectively).

Aircel Ln-transformed descriptive statistics					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
lwoutd	521	.750414	1.057799	0	7.521318
lwinde	517	.7526602	1.059848	0	7.365813
lclus	337	-1.051558	.3484488	-3.516868	-.5389965
leige	517	-10.19112	1.998187	-11.82764	0
Key lclus = Ln(Clustering Coefficient). leige = Ln(Eigenvector Centrality). lwoutd = Ln(Weighted Out-Degree). lwinde = Ln(Weighted In-Degree).					

Table 5-10: Aircel Ln-transformed descriptive statistics.

Bharti Airtel Ln-transformed descriptive statistics					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
lwoutd	179	.6897406	.9915971	0	5.273
lwinde	177	.6939203	1.000111	0	4.997212
lclus	111	-1.087415	.4595142	-3.484855	-.6931472
leige	177	-6.338549	1.98078	-8.963168	0
Key lclus = Ln(Clustering Coefficient). leige = Ln(Eigenvector Centrality). lwoutd = Ln(Weighted Out-Degree). lwinde = Ln(Weighted In-Degree).					

Table 5-11: Bharti Airtel Ln-transformed descriptive statistics.

Vodafone Ln-transformed descriptive statistics					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
lwoutd	1512	.7441494	1.099229	0	9.340842
lwinde	1513	.744936	1.100011	0	9.27669
lclus	1141	-.9500353	.3362528	-3.401197	-.4054651
leige	1513	-12.54449	1.528162	-13.70155	0
Key lclus = Ln(Clustering Coefficient). leige = Ln(Eigenvector Centrality). lwoutd = Ln(Weighted Out-Degree). lwinde = Ln(Weighted In-Degree).					

Table 5-12: Vodafone Ln-transformed descriptive statistics.

Based on the differences of the *Descriptive Statistics* between the three operator networks, one would expect that the level of hierarchical network structuring would differ for each mobile broadband operator network. Especially the *Eigenvector Centrality*, which covers vertex influences in a network, is considerably higher for the Bharti Airtel observations (see Table 5-8 above). It therefore seems, that the hierarchical organisation of the three networks should be more apparent for Aircel and Vodafone, compared to Bharti Airtel. To summarize, the separation of the complete dataset into operator-based models seems to be the most fruitful and pragmatic way for exploring the network structural differences between the three Tamil Nadu mobile broadband operators. This furthermore allows us to compare the underlying differences between the three upstream connectivity networks in this case study.

5.3.2 Econometric Models

This section introduces four different econometric models to test the *Working Hypotheses* discussed in the Literature Review in section 2.5. We start by comparing four alternative functional form specifications (linear-linear, log-linear, linear-log and log-log), for two separate models. The first one tests the relationship between *Out-Degree* connectivity, and level of hierarchical structuring (represented by the *Clustering Coefficient* and the *Eigenvector Centrality*). The second focusses on the relation between *In-Degree* connectivity and the level of hierarchical structuring (*Clustering Coefficient* and

Eigenvector Centrality). The initial estimations are conducted on the full dataset, unfiltered by mobile broadband operator of origin, exclusively to explore and compare the alternative functional forms. Next, in the following sections the full analysis will be done separately for each mobile broadband operator of origin. The four functional form specifications for the *Out-Degree* connectivity and level of hierarchical structuring (*Clustering Coefficient* and *Eigenvector Centrality*) relationship are:

Linear-linear:

$$\begin{aligned} woutd_{All\ operators} = \\ \beta_0 + \beta_1(clus_{All\ operators}) + \beta_2(eige_{All\ operators}) + \varepsilon \end{aligned}$$

Log-linear:

$$\begin{aligned} \ln(woutd)_{All\ operators} = \\ \beta_0 + \beta_1(clus_{All\ operators}) + \beta_2(eige_{All\ operators}) + \varepsilon \end{aligned}$$

Linear-log:

$$\begin{aligned} woutd_{All\ operators} = \\ \beta_0 + \beta_1 \ln(clus_{All\ operators}) + \beta_2 \ln(eige_{All\ operators}) + \varepsilon \end{aligned}$$

Log-log:

$$\begin{aligned} \ln(woutd)_{All\ operators} = \\ \beta_0 + \beta_1 \ln(clus_{All\ operators}) + \beta_2 \ln(eige_{All\ operators}) + \varepsilon \end{aligned}$$

The four functional form specifications for the *In-Degree* connectivity and Clustering (*Clustering Coefficient* and *Eigenvector Centrality*) relationships are given by:

Linear-linear:

$$\begin{aligned} winde_{All\ operators} = \\ \beta_0 + \beta_1(clus_{All\ operators}) + \beta_2(eige_{All\ operators}) + \varepsilon \end{aligned}$$

Log-linear:

$$\begin{aligned} \ln(winde)_{All\ operators} = \\ \beta_0 + \beta_1 \ln(clus_{All\ operators}) + \beta_2 \ln(eige_{All\ operators}) + \varepsilon \end{aligned}$$

Linear-log:

$$\begin{aligned} winde_{All\ operators} = \\ \beta_0 + \beta_1 \ln(clus_{All\ operators}) + \beta_2 \ln(eige_{All\ operators}) + \varepsilon \end{aligned}$$

Log-log:

$$\begin{aligned} \ln(winde_{All\ operators}) = \\ \beta_0 + \beta_1 \ln(clus_{All\ operators}) + \beta_2 \ln(eige_{All\ operators}) + \varepsilon \end{aligned}$$

Estimation results for alternative functional form specification (<i>Out-Degree</i> connectivity and Hierarchical Structuring (<i>Clustering Coefficient</i> and <i>Eigenvector Centrality</i>))	
Linear-Linear Dependent variable: woutd Covariates: clus, eig $F(2, 2212) = 657.41$ $\text{Prob} > F = 0.000$ $R\text{-Squared} = 0.3728$ $\text{Root MSE} = 218.12$	Log-Linear Dependent variable: lwoutd Covariates: clus, eig $F(2, 2209) = 112.62$ $\text{Prob} > F = 0.000$ $R\text{-Squared} = 0.0933$ $\text{Root MSE} = 1.0297$
Linear-Log Dependent variable: woutd Covariates: lclus, leig $F(2, 1585) = 42.04$ $\text{Prob} > F = 0.000$ $R\text{-Squared} = 0.0504$ $\text{Root MSE} = 316.86$	Log-Log * Dependent variable: lwoutd Covariates: lclus, leige $F(2, 1582) = 847.43$ $\text{Prob} > F = 0.000$ $R\text{-Squared} = 0.5172$ $\text{Root MSE} = .80506$
Key * chosen functional form. clus = Clustering Coefficient. eig = Eigenvector Centrality. woutd = Weighted Out-Degree. winde = Weighted In-Degree. lclus = Ln(Clustering Coefficient). leige = Ln(Eigenvector Centrality). lwoutd = Ln(Weighted Out-Degree). lwinde = Ln(Weighted In-Degree).	

Table 5-13: Functional form specification Model 1.

Estimation results for alternative functional form specification (<i>In-Degree connectivity and Hierarchical Structuring (Clustering Coefficient and Eigenvector Centrality)</i>)	
Linear-Linear Dependent variable: winde Covariates: clus, eig $F(2, 2212) = 624.70$ Prob > F = 0.000 R-Squared = 0.3610 Root MSE = 211.42	Log-Linear Dependent variable: lwinde Covariates: clus, eig $F(2, 2204) = 119.52$ Prob > F = 0.000 R-Squared = 0.0978 Root MSE = 1.0288
Linear-Log Dependent variable: winde Covariates: lclus, leig $F(2, 1585) = 52.25$ Prob > F = 0.000 R-Squared = 0.0619 Root MSE = 302.42	Log-Log * Dependent variable: lwinde Covariates: lclus, leig $F(2, 1585) = 902.88$ Prob > F = 0.0000 R-Squared = 0.5326 Root MSE = .79525
Key * chosen functional form. clus = Clustering Coefficient. eig = Eigenvector Centrality. woutd = Weighted Out-Degree. winde = Weighted In-Degree. lclus = Ln(Clustering Coefficient). leig = Ln(Eigenvector Centrality). lwoutd = Ln(Weighted Out-Degree). lwinde = Ln(Weighted In-Degree).	

Table 5-14: Functional form specification Model 2.

Based on these preliminary estimations using the full operator dataset (comprising the full observations of all mobile broadband operators) and by considering the Degrees of Freedom and the (Adjusted) R-Squares, we decide to proceed with the log-log specification in the following *Statistical Network Analysis* for the three-separate mobile broadband operators. The parameters of the variable coefficients for the given estimations above were not reported here, as they will be discussed for the operator-specific models below. Given the preliminary functional form estimation, in this section we develop several econometric models using the mobile broadband operator-based subsets of the collected dataset.

In the first stage of the estimation we consider two different models, Model 1 with the Network Structural Markers (*Clustering Coefficient* and *Eigenvector Centrality*) as covariates and the outgoing upstream connectivity (*Weighted Out-Degree*) as dependent variable.

Model 1:

$$\ln(woutd_{per\ operator}) = \beta_0 + \beta_1 \ln(clus_{per\ operator}) + \beta_2 \ln(eige_{per\ operator}) + \varepsilon$$

Next, Model 2 is estimated using again the Network Structural Markers (*Clustering Coefficient* and *Eigenvector Centrality*) as covariates but here the incoming upstream connectivity (*Weighted In-Degree*) as dependent variable.

Model 2:

$$\ln(winde_{per\ operator}) = \beta_0 + \beta_1 \ln(clus_{per\ operator}) + \beta_2 \ln(eige_{per\ operator}) + \varepsilon$$

In the following section, we start with the specification and the estimation of Model 1, followed by Model 2 in section 5.3.4.

5.3.3 Model 1 (Weighted Out-Degree connectivity)

Following the Network Analysis at Autonomous System granularity in Chapter 4, the elaboration of the first stage, of the two-stage econometric model, is separately performed, first based on the Aircel observations only, then on the Bharti Airtel observations, and lastly on the Vodafone ones. As described in section 5.3.2, our first econometric model (Model 1) in stage one is estimated using a log-log functional form specification for the *Weighted Out-Degree* connectivity and Clustering (*Clustering Coefficient* and *Eigenvector Centrality*) relationships. The estimations for each mobile broadband operator are given by:

Aircel Model 1:

$$\ln(woutd_{Aircel}) = \beta_0 + \beta_1 \ln(clus_{Aircel}) + \beta_2 \ln(eige_{Aircel}) + \varepsilon$$

Bharti Airtel Model 1:

$$\ln(woutd_{Bharti\ Airtel}) = \beta_0 + \beta_1 \ln(clus_{Bharti\ Airtel}) + \beta_2 \ln(eige_{Bharti\ Airtel}) + \varepsilon$$

Vodafone Model 1:

$$\ln(woutd_{Vodafone}) = \beta_0 + \beta_1 \ln(clus_{Vodafone}) + \beta_2 \ln(eige_{Vodafone}) + \varepsilon$$

Below we start with the specification and estimation of Model 1 for the Aircel observations, followed by the Bharti Airtel and the Vodafone observations before summarising the key findings.

Aircel Model 1

The dependent variable, *lwoutd*, and the covariates of interest *lclus* and *leige*, are continuous. Table 5-15 below reports the estimates for Aircel Model 1 based on a robust estimation method (Huber / White / Sandwich Std. Errors) for the adopted log-log specification (see section 5.3.3 above), using 336 vertex observations, obtained from the Aircel graph, G_{Aircel_AS} , as elaborated in Chapter 4.

The robust estimation techniques were required given the presence of heteroskedasticity that was observed in a preliminary OLS estimation not reported here. From Table 5-15 below, we can see that the overall model is significant (p-value of 0.0000 for $F(2, 333) = 114.05$) and explains 66.36% (R-Squared equals 0.6636) of the total variation of the $\text{Ln}(\text{Weighted Out-Degree})$.

Model 1 Multiple Linear regression – Aircel Huber / White / Sandwich Std. Errors						
Variable abbreviations <i>lclus</i> : Ln(Clustering Coefficient) <i>leige</i> : Ln(Eigenvector Centrality) <i>lwoutd</i> : Ln(Weighted Out-Degree)				Number of obs = 336 $F(2, 333) = 114.05$ Prob > F = 0.0000 R-squared = 0.6636 Root MSE = .65508		
<i>lwoutd</i>	Coef.	Std. Error	t	p > t	[95% Conf. Interval]	
<i>lclus</i>	-1.182964***	.2442479	-4.84	0.000	-1.663427	-.7025002
<i>leige</i>	.3212572***	.0344788	9.32	0.000	.2534335	.389081
_cons	2.98095***	.5150569	5.79	0.000	1.967775	3.994126
Key * p < .05; ** p < .01; *** p < .001						

Table 5-15: Aircel multiple robust log-log regression Model 1.

The t-tests showed that the two key covariates were statistically significant at 99% level. The coefficients for the main covariates indicate that *lclus* had a negative effect on the *lwoutd* (-1.182964***), whereas *leige* showed a positive effect on *lwoutd* (.3212572***).

Hence, from these estimates, we can see that a 1% increase of *lclus* represents a decrease in *Weighted Out-Degree* connectivity (*lwoutd*) by 1.18%. Similarly, a 1% increase of *leige* represents an increase in *Weighted Out-Degree* connectivity (*lwoutd*) by 0.32%.

The Ramsey RESET test for $F(3, 330) = 3.41$, a p-value of 0.0178, revealed an Omitted Variable Bias, at 95%. However, having carried additional estimations for more functional form specifications including higher power of the explanatory variables up to the fourth order (not reported here), we saw no improvement in this test, hence we retained the original log-log specification. The estimated Variance Inflation Factor (VIF) was below the thresholds that would indicate the presence of multicollinearity. The plot of the predicted residual distributions, see Figure 5-16 below, shows an approximately normal distribution ranging closely around the mean of zero.

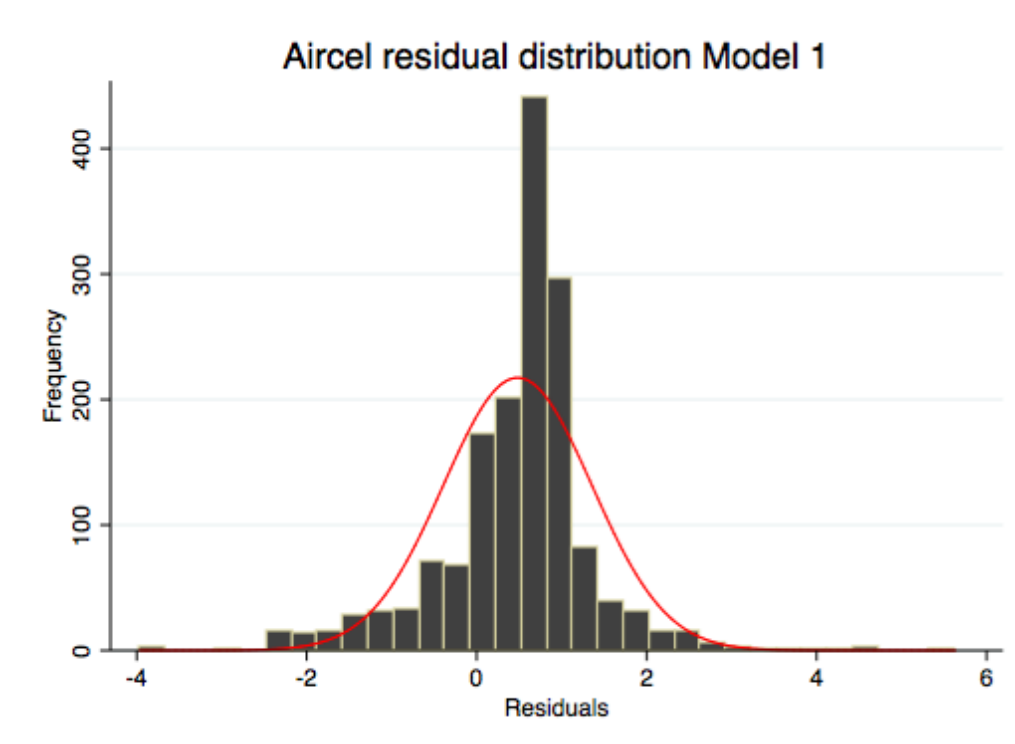


Figure 5-16: Aircel predicted error term residual distribution Model 1.

Bharti Airtel Model 1

Again, the dependent variable, *lwoutd*, and the covariates of interest *lclus* and *leige*, are continuous. Table 5-16 below reports the estimates for Bharti Airtel Model 1 based on a robust estimation method (Huber / White / Sandwich Std. Errors) for the adopted log-log specification (see section 5.3.3 above), using 109 vertex observations, obtained from the Bharti Airtel graph, $G_{Bharti\ Airtel_AS}$, as elaborated in Chapter 4.

The robust estimation techniques were required given the presence of heteroskedasticity that was observed in a preliminary OLS estimation not reported here. From Table 5-16 below, we can see that the overall model is significant (p-value of 0.0000 for $F(2, 106) =$

15.22) and explains 48.43% (R-Squared equals 0.4843) of the total variation of the $\text{Ln}(\text{Weighted Out-Degree})$.

The t-tests showed that the two key covariates were statistically significant at 95% level for *lclus* and the 99% level for *leige*. The coefficients for the main covariates indicate that *lclus* had a negative effect on the *lwoutd* (-.9986846**), whereas *leige* showed a positive effect on *lwoutd* (.277029***).

Model 1 Multiple Linear regression – Bharti Airtel Huber / White / Sandwich Std. Errors						
Variable abbreviations <i>lclus</i> : Ln(Clustering Coefficient) <i>leige</i> : Ln(Eigenvector Centrality) <i>lwoutd</i> : Ln(Weighted Out-Degree)				Number of obs = 109 F(2, 106) = 15.22 Prob > F = 0.0000 R-squared = 0.4843 Root MSE = .72944		
<i>lwoutd</i>	Coef.	Std. Error	t	p > t	[95% Conf. Interval]	
<i>lclus</i>	-.9986846**	.3569365	-2.80	0.006	-1.706346	-.2910233
<i>leige</i>	.277029***	.0631923	4.38	0.000	.151744	.4023139
_cons	1.561349**	.5604904	2.79	0.006	.4501226	2.672576
Key * p < .05; ** p < .01; *** p < .001						

Table 5-16: Bharti Airtel multiple robust log-log regression Model 1.

Hence, from these estimates, we can see that a 1% increase of *lclus* represents a decrease in *Weighted Out-Degree* connectivity (*lwoutd*) by 0.99%. Similarly, a 1% increase of *leige* represents an increase in *Weighted Out-Degree* connectivity (*lwoutd*) by 0.28%.

The Ramsey RESET test for $F(3, 103) = 16.94$, a p-value of 0.0000, revealed an Omitted Variable Bias, at 99%. However, having carried additional estimations for more functional form specifications including higher power of the explanatory variables up to the fourth order (not reported here), we saw no improvement in this test, hence we again retained the original log-log specification. The estimated Variance Inflation Factor (VIF) was below the thresholds that would indicate the presence of multicollinearity. The plot

of the predicted residual distributions, see Figure 5-16 below, shows an approximately normal distribution ranging closely around the mean of zero.

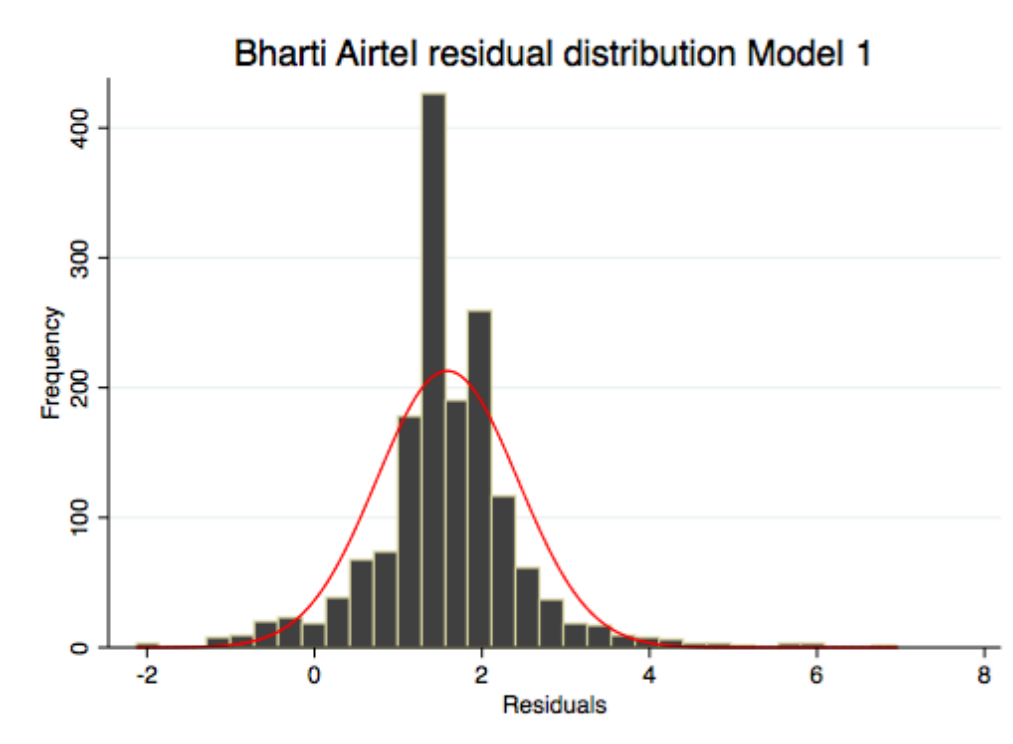


Figure 5-17: Bharti Airtel predicted error term residual distribution Model 1.

Vodafone Model 1

Like previously, the dependent variable, *lwoutd*, and the covariates of interest *lclus* and *leige*, are continuous. Table 5-17 below reports the estimates for Vodafone Model 1 based on a robust estimation method (Huber / White / Sandwich Std. Errors) for the adopted log-log specification (see section 5.3.3 above), using 1140 vertex observations, obtained from the Vodafone graph, $G_{Vodafone_AS}$, as elaborated in Chapter 4.

Again, the robust estimation techniques were required given the presence of heteroskedasticity that was observed in a preliminary OLS estimation not reported here. From Table 5-17 below, we can see that the overall model is significant (p-value of 0.0000 for $F(2, 1137) = 751.76$) and explains 79.15% (R-Squared equals 0.7915) of the total variation of the $\text{Ln}(\text{Weighted Out-Degree})$.

Model 1 Multiple Linear regression – Vodafone Huber / White / Sandwich Std. Errors						
Variable abbreviations lclus: Ln(Clustering Coefficient) leige: Ln(Eigenvector Centrality) lwoutd: Ln(Weighted Out-Degree)				Number of obs = 1140 F(2, 1137) = 751.76 Prob > F = 0.0000 R-squared = 0.7915 Root MSE = .53825		
lwoutd	Coef.	Std. Error	t	p > t	[95% Conf. Interval]	
lclus	-.8647558***	.1388745	-6.23	0.000	-1.137235	-.5922768
leige	.5424204***	.0236467	22.94	0.000	.4960243	.5888165
_cons	6.794033***	.3973645	17.10	0.000	6.014383	7.573683
Key * p < .05; ** p < .01; *** p < .001						

Table 5-17: Vodafone multiple robust log-log regression Model 1.

The t-tests showed that the two key covariates were statistically significant at 99% level. The coefficients for the main covariates indicate that *lclus* had a negative effect on the *lwoutd* (-.8647558**), whereas *leige* showed a positive effect on *lwoutd* (.5424204***).

Hence, from these estimates, we can see that a 1% increase of *lclus* represents a decrease in *Weighted Out-Degree* connectivity (*lwoutd*) by 0.86%. Similarly, a 1% increase of *leige* represents an increase in *Weighted Out-Degree* connectivity (*lwoutd*) by 0.54%.

The Ramsey RESET test for $F(3, 103) = 1.25$, a p-value of 0.2921, revealed no Omitted Variable Bias. Hence, we retained the original log-log specification. Interestingly, dropping the China Education and Research Network (AS4538) outlier observation (with the highest *Weighted Out-Degree*) from the dataset would lead to a lower Omitted Variable Bias at 90% level. The estimated Variance Inflation Factor (VIF) was below the thresholds that would indicate the presence of multicollinearity. The plot of the predicted residual distributions, see Figure 5-18 below, shows an approximately normal distribution ranging closely around the mean of zero.

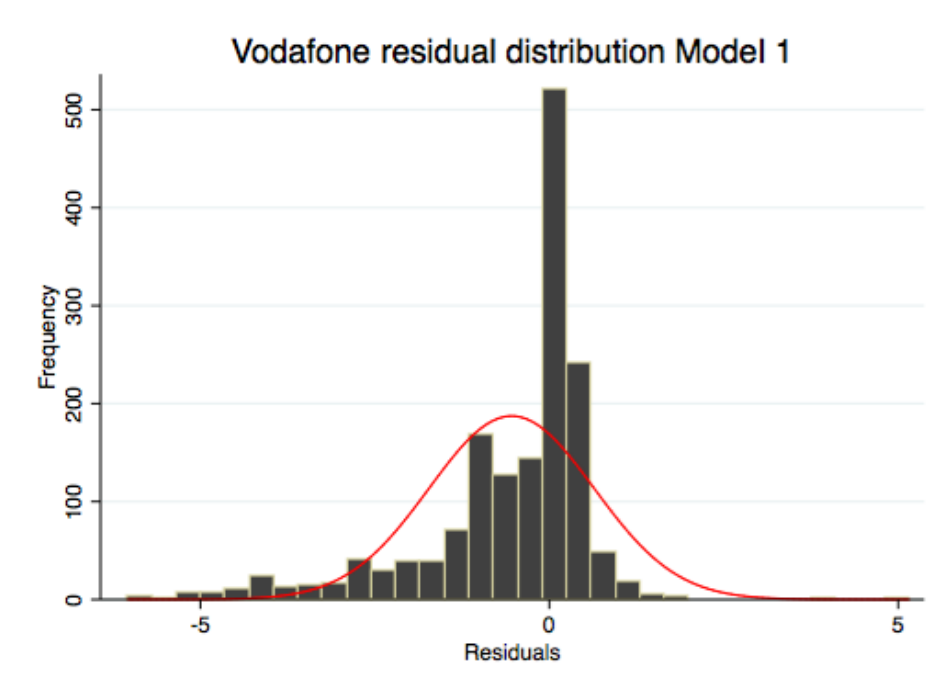


Figure 5-18: Vodafone predicted error term residual distribution Model 1.

Summary Model 1

Given the slope of the log-log regression between the *Weighted Out-Degree* and the *Clustering Coefficient*, we can see that an **increase** in $\text{Ln}(\text{Clustering Coefficient})$ is associated with a **decrease** in outgoing connectivity (here represented by the variable of the $\text{Ln}(\text{Weighted Out-Degree})$).

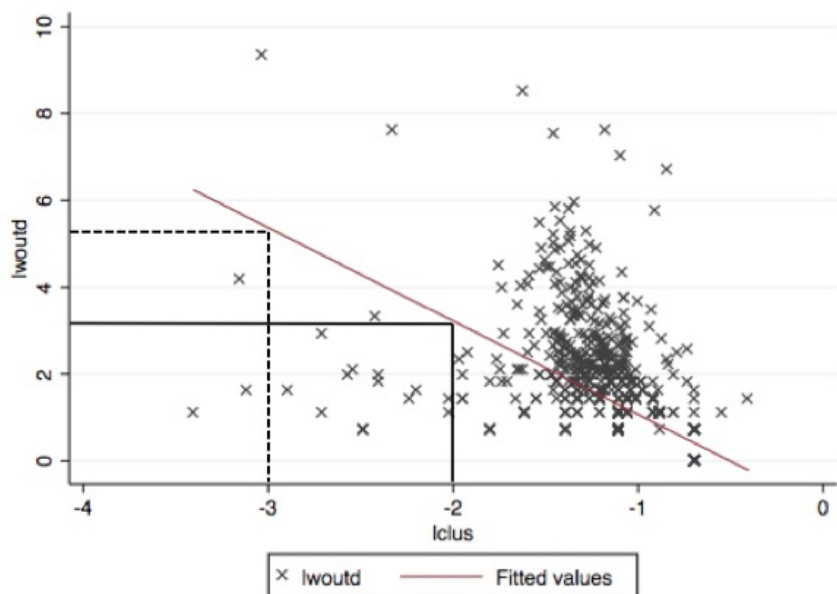


Figure 5-19: Predicted effects of an increase in $\text{Ln}(\text{Clustering Coefficient})$ in $\text{Ln}(\text{Weighted Out-Degree})$ - Model 1.

Second, given the slope of the log-log regression between the *Weighted Out-Degree* and the *Eigenvector*, we indicate that an **increase** in $\text{Ln}(\text{Eigenvector Centrality})$ is associated with an **increase** in outgoing connectivity (again captured by the $\text{Ln}(\text{Weighted Out-Degree})$).

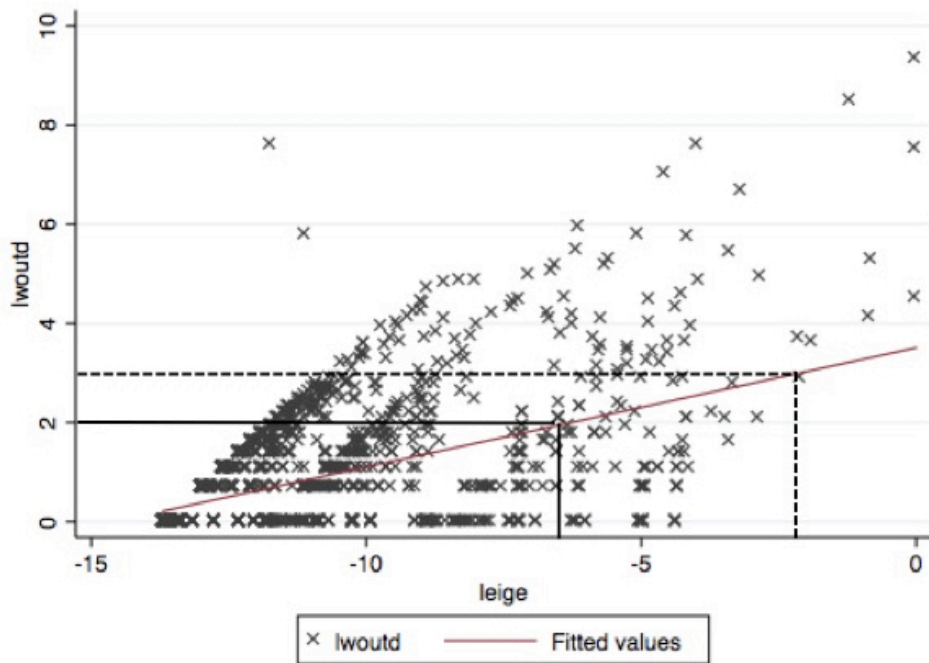


Figure 5-20: Predicted effects of an increase in $\text{Ln}(\text{Eigenvector Centrality})$ in $\text{Ln}(\text{Weighted Out-Degree})$ - Model 1.

5.3.4 Model 2 (Weighted In-Degree)

Similarly to the estimations of Model 1 above, we assume that studying the *Weighted In-Degree* connectivity, in this Model 2, should also reveal statistically significant negative *lclus* and positive *leige* coefficients, when robustly being regressed against *lwinde*. Again, following the Network Analysis at Autonomous System granularity in Chapter 4, the specification and estimation of the econometric models is separately performed. Again, as described in section 5.3.2 above, our second econometric model (Model 2) in stage one is estimated using a log-log functional form specification for the *Weighted In-Degree* connectivity and hierarchical structuring (*Clustering Coefficient* and *Eigenvector Centrality*) relationships. The estimations for each mobile broadband operator are given by:

Aircel Model 2:

$$\ln(winde_{Aircel}) = \beta_0 + \beta_1 \ln(clus_{Aircel}) + \beta_2 \ln(eige_{Aircel}) + \varepsilon$$

Bharti Airtel Model 2:

$$\ln(winde_{Bharti Airtel}) = \beta_0 + \beta_1 \ln(clus_{Bharti Airtel}) + \beta_2 \ln(eige_{Bharti Airtel}) + \varepsilon$$

Vodafone Model 2:

$$\ln(winde_{Vodafone}) = \beta_0 + \beta_1 \ln(clus_{Vodafone}) + \beta_2 \ln(eige_{Vodafone}) + \varepsilon$$

In the section below, we start with the specification and estimation of Model 2 for the Aircel observations, followed by the Bharti Airtel observations and the Vodafone ones before again summarising the key findings.

Aircel Model 2

The dependent variable, $lwinde$, and the covariates of interest $lclus$ and $leige$, are continuous. Table 5-18 below reports the estimates for Aircel Model 2 based on a robust estimation method (Huber / White / Sandwich Std. Errors) for the adopted log-log specification (see section 5.3.3 above), using 337 vertex observations, obtained from the Aircel graph, G_{Aircel_AS} , as elaborated in Chapter 4.

Again, the robust estimation techniques were required given the presence of heteroskedasticity that was observed in a preliminary OLS estimation not reported here. From Table 5-18, we can see that the overall model is significant (p-value of 0.0000 for $F(2, 334) = 145.25$) and explains 72.18% (R-Squared equals 0.7218) of the total variation of the $\ln(Weighted In-Degree)$.

The t-tests showed that the two key covariates were statistically significant at 99% level.

The coefficients for the main covariates indicate that *lclus* had a negative effect on the *lwinde* (-1.104983***), whereas *leige* showed a positive effect on *lwinde* (.334886***).

Model 2 Multiple Linear regression – Aircel Huber / White / Sandwich Std. Errors						
Variable abbreviations <i>lclus</i> : Ln(Clustering Coefficient) <i>leige</i> : Ln(Eigenvector Centrality) <i>lwinde</i> : Ln(Weighted In-Degree)				Number of obs = 337 F(2, 334) = 145.25 Prob > F = 0.0000 R-squared = 0.7218 Root MSE = .59678		
<i>lwinde</i>	Coef.	Std. Error	t	p > t	[95% Conf. Interval]	
<i>lclus</i>	-1.104983***	.1919853	-5.76	0.000	-1.482636	-.7273301
<i>leige</i>	.334886***	.0306652	10.92	0.000	.2745648	.3952073
_cons	3.169333***	.4391155	7.22	0.000	2.305552	4.033114
Key * p < .05; ** p < .01; *** p < .001						

Table 5-18: Aircel multiple robust log-log regression Model 2.

Hence, from these estimates, we can see that a 1% increase of *lclus* represents a decrease in *Weighted In-Degree* connectivity (*lwinde*) by 1.10%. Similarly, a 1% increase of *leige* represents an increase in *Weighted In-Degree* connectivity (*lwinde*) by 0.33%. The Ramsey RESET test for $F(3, 331) = 6.04$, a p-value of 0.0005, revealed an Omitted Variable Bias, at 99%. However, having carried additional estimations for more functional form specifications including higher power of the explanatory variables up to the fourth order (not reported here), we saw no improvement in this test, hence we retained the original log-log specification. The estimated Variance Inflation Factor (VIF) was below the thresholds that would indicate the presence of multicollinearity. The plot of the predicted residual distributions, see Figure 5-21 below, shows an approximately normal distribution ranging closely around the mean of zero.

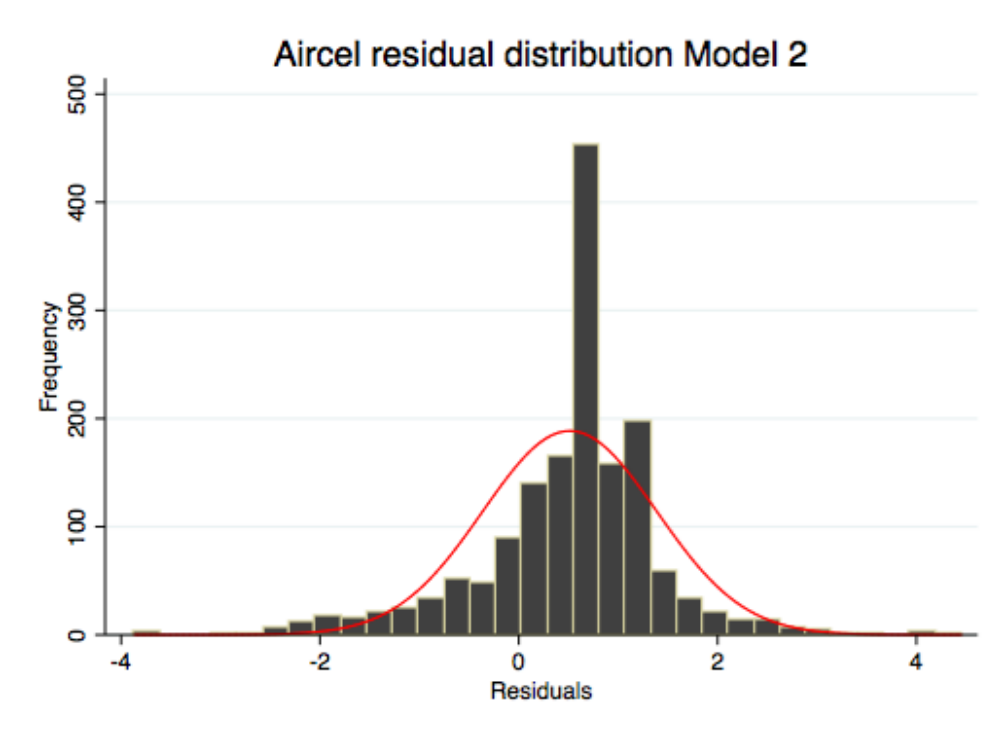


Figure 5-21: Airtel predicted error term residual distribution Model 2.

Bharti Airtel Model 2

Again, the dependent variable, *lwinde*, and the covariates of interest *lclus* and *leige*, are continuous. Table 5-19 below reports the estimates for Bharti Airtel Model 2 based on a robust estimation method (Huber / White / Sandwich Std. Errors) for the adopted log-log specification (see section 5.3.3 above), using 110 vertex observations, obtained from the Bharti Airtel graph, $G_{Bharti\ Airtel_AS}$, as elaborated in Chapter 4.

The robust estimation techniques were required given the presence of heteroskedasticity that was observed in a preliminary OLS estimation not reported here. From Table 5-19 below, we can see that the overall model is significant (p-value of 0.0000 for $F(2, 107) = 23.04$) and explains 54.35% (R-Squared equals 0.5435) of the total variation of the $\text{Ln}(\text{Weighted In-Degree})$.

The t-tests showed that the two key covariates were statistically significant at 99% level. The coefficients for the main covariates indicate that *lclus* had a negative effect on the *lwinde* (-1.032788***), whereas *leige* showed a positive effect on *lwinde* (.2765935***).

Model 2 Multiple Linear regression – Bharti Airtel Huber / White / Sandwich Std. Errors						
Variable abbreviations lclus: Ln(Clustering Coefficient) leige: Ln(Eigenvector Centrality) lwinde: Ln(Weighted In-Degree)				Number of obs = 110 F(2, 107) = 23.04 Prob > F = 0.0000 R-squared = 0.5435 Root MSE = .72773		
lwinde	Coef.	Std. Error	t	p > t	[95% Conf. Interval]	
lclus	-1.032788***	.2813437	-3.67	0.000	-1.590519	-.4750567
leige	.2765935***	.0612059	4.52	0.000	.15526	.397927
_cons	1.51585**	.5292865	2.86	0.005	.4666014	2.565099
Key * p < .05; ** p < .01; *** p < .001						

Table 5-19: Bharti Airtel multiple robust log-log regression Model 2.

Hence, from these estimates, we can see that a 1% increase of *lclus* represents a decrease in *Weighted In-Degree* connectivity (*lwinde*) by 1.03%. Similarly, a 1% increase of *leige* represents an increase in *Weighted In-Degree* connectivity (*lwinde*) by 0.28%.

The Ramsey RESET test for $F(3, 104) = 8.75$, a p-value of 0.0000, revealed an Omitted Variable Bias, at 99%. However, having carried additional estimations for more functional form specifications including higher power of the explanatory variables up to the fourth order (not reported here), we saw no improvement in this test, hence we again retained the original log-log specification. The estimated Variance Inflation Factor (VIF) was below the thresholds that would indicate the presence of multicollinearity. The plot of the predicted residual distributions, see Figure 5-22 below, shows an approximately normal distribution ranging closely around the mean of zero.

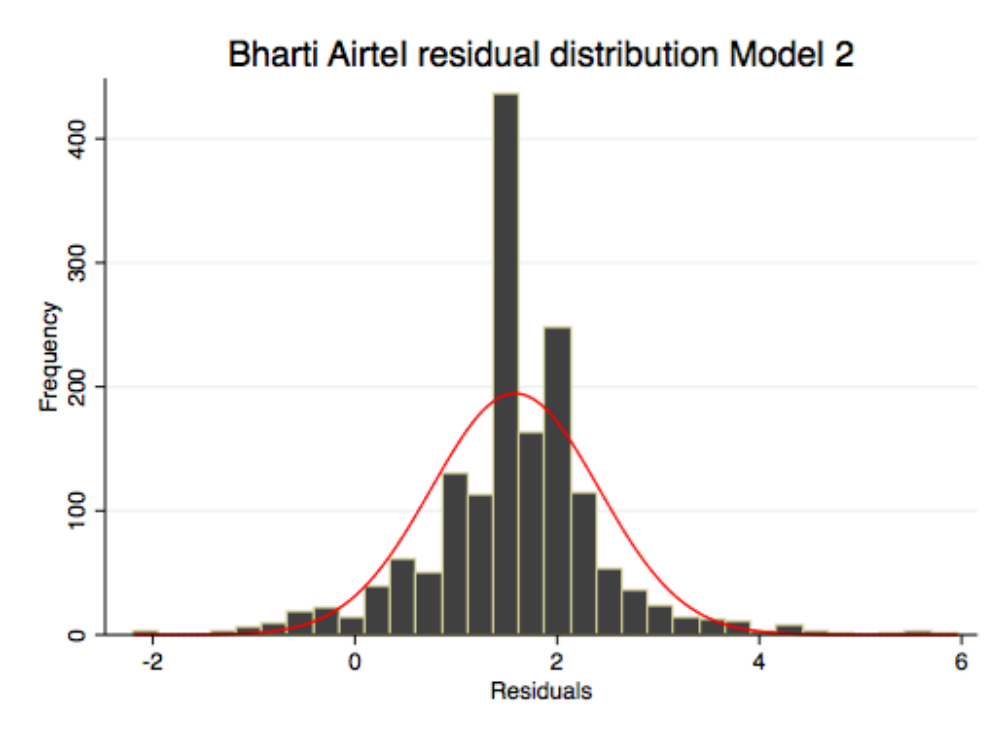


Figure 5-22: Bharti Airtel predicted error term residual distribution Model 2.

Vodafone Model 2

Like previously, the dependent variable, *lwinde*, and the covariates of interest *lclus* and *leige*, are continuous. Table 5-20 below reports the estimates for Vodafone Model 2 based on a robust estimation method (Huber / White / Sandwich Std. Errors) for the adopted log-log specification (see section 5.3.3 above), using 1141 vertex observations, obtained from the Vodafone graph, $G_{Vodafone_AS}$, as elaborated in Chapter 4.

Again, the robust estimation techniques were required given the presence of heteroskedasticity that was observed in a preliminary OLS estimation not reported here. From Table 5-20 below, we can see that the overall model is significant (p-value of 0.0000 for $F(2, 1138) = 859.71$) and explains 80.88% (R-Squared equals 0.8088) of the total variation of the $\text{Ln}(\text{Weighted In-Degree})$.

The t-tests showed that the two key covariates were statistically significant at 99% level. The coefficients for the main covariates indicate that *lclus* had a negative effect on the *lwinde* (-.7859873***), whereas *leige* showed a positive effect on *lwinde* (.5514546***).

Model 2 Multiple Linear regression – Vodafone Huber / White / Sandwich Std. Errors						
Variable abbreviations lclus: Ln(Clustering Coefficient) leige: Ln(Eigenvector Centrality) lwinde: Ln(Weighted In-Degree)				Number of obs = 1141 F(2, 1138) = 859.71 Prob > F = 0.0000 R-squared = 0.8088 Root MSE = .51578		
lwinde	Coef.	Std. Error	t	p > t	[95% Conf. Interval]	
lclus	-.7859873***	.10973	-7.16	0.000	-1.001283	-.5706915
leige	.5514546***	.0218291	25.26	0.000	.5086248	.5942843
_cons	6.974716***	.3532901	19.74	0.000	6.281543	7.667889
Key * p < .05; ** p < .01; *** p < .001						

Table 5-20: Vodafone multiple robust log-log regression Model 2.

Hence, from these estimates, we can see that a 1% increase of *lclus* represents a decrease in *Weighted In-Degree* connectivity (*lwinde*) by 0.78%. Similarly, a 1% increase of *leige* represents an increase in *Weighted In-Degree* connectivity (*lwinde*) by 0.55%.

The Ramsey RESET test for $F(3, 1135) = 2.62$, a p-value of 0.0497, revealed an Omitted Variable Bias, at 90%. However, having carried additional estimations for more functional form specifications including higher power of the explanatory variables up to the fourth order (not reported here), we saw no improvement in this test, hence we again retained the original log-log specification.

The estimated Variance Inflation Factor (VIF) was below the thresholds that would indicate the presence of multicollinearity. The plot of the predicted residual distributions, see Figure 5-23 below, shows an approximately normal distribution ranging closely around the mean of zero.

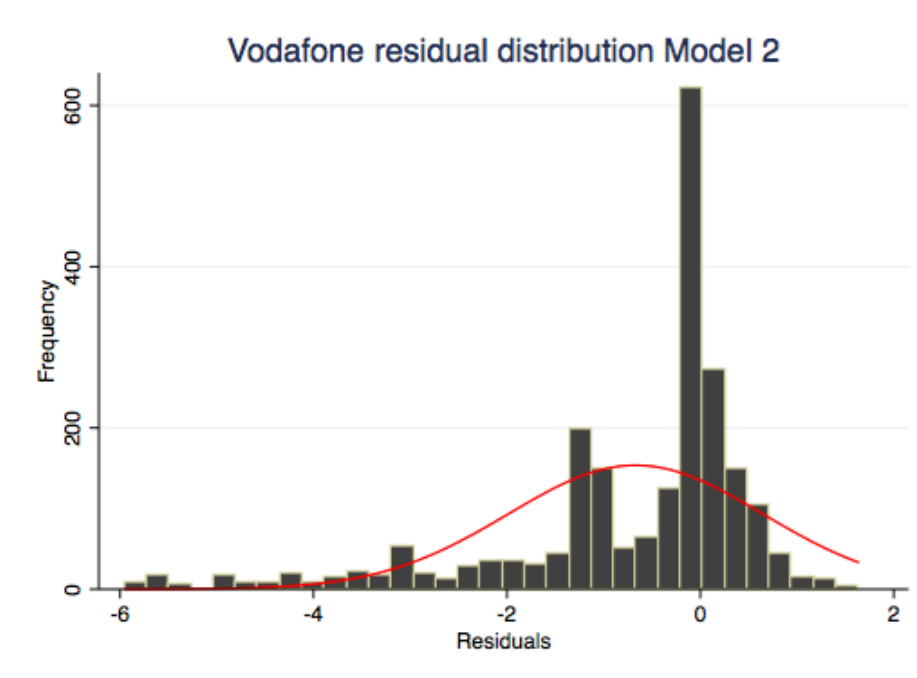


Figure 5-23: Vodafone predicted error term residual distribution Model 2.

Summary Model 2

Given the slope of the log-log regression between the *Weighted In-Degree* and the *Clustering Coefficient*, see Figure 5-24 below, we can see that an **increase** in $\text{Ln}(\text{Clustering Coefficient})$, is associated with a **decrease** in incoming connectivity (here represented by the variable of the $\text{Ln}(\text{Weighted In-Degree})$).

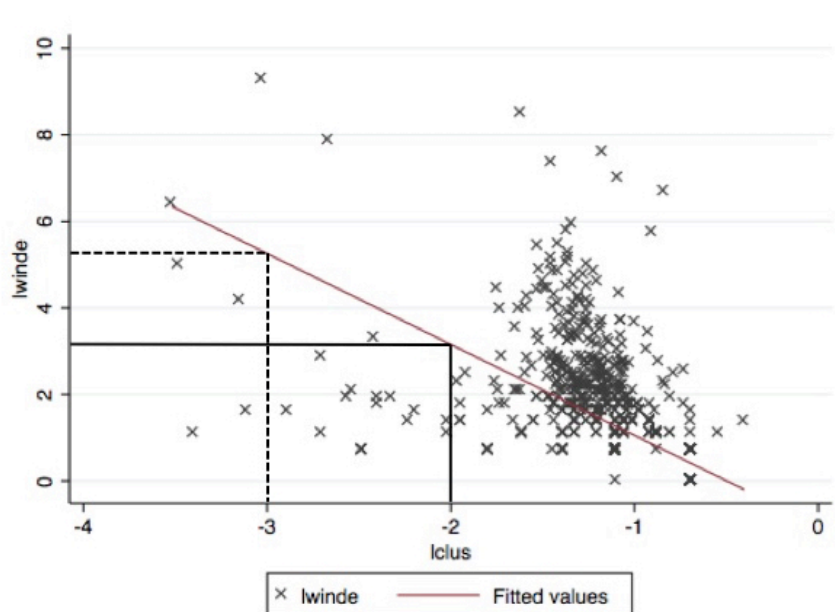


Figure 5-24: Predicted effects of an increase in $\text{Ln}(\text{Clustering Coefficient})$ on $\text{Ln}(\text{Weighted In-Degree})$ - Model 2.

Second, given the slope of the log-log regression between the *Weighted In-Degree* and the *Eigenvector Centrality*, see Figure 5-25 below, we also indicate that an **increase** in $\text{Ln}(\text{Eigenvector Centrality})$ is associated with an **increase** in incoming connectivity (again captured by the $\text{Ln}(\text{Weighted In-Degree})$).

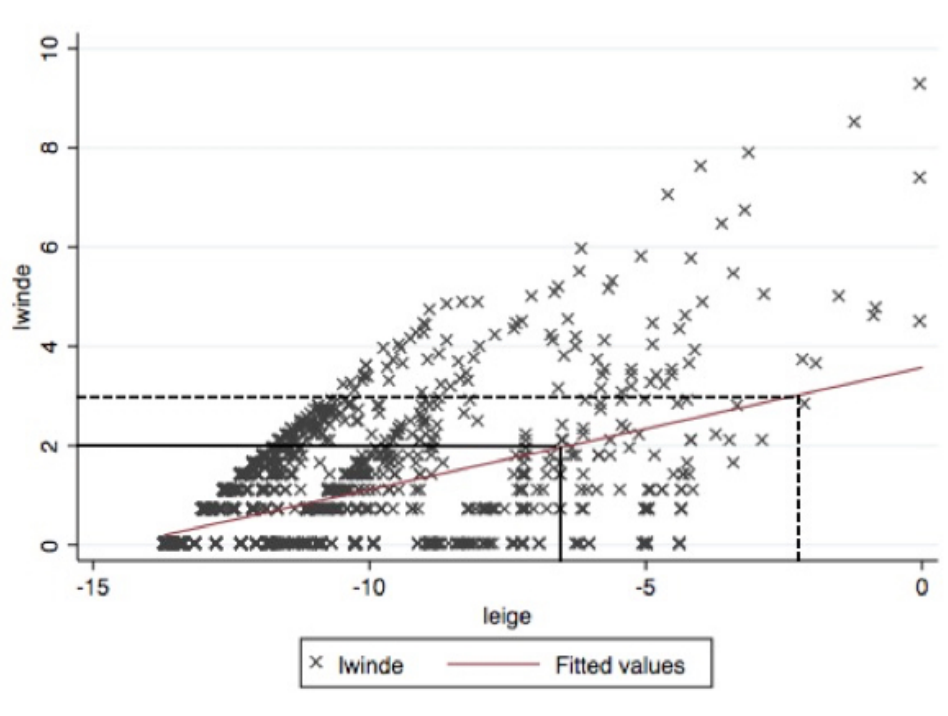


Figure 5-25: Predicted effects of an increase in $\text{Ln}(\text{Eigenvector Centrality})$ on $\text{Ln}(\text{Weighted In-Degree})$ - Model 2.

The findings of Model 2 (based on the *Weighted In-Degree* connectivity) support the estimations of Model 1 (based on the *Weighted Out-Degree* connectivity). Therefore, we conclude that the *Clustering Coefficient* and the *Eigenvector Centrality* relate to mobile broadband operator connectivity and hierarchical structuring (measured through either the *Weighted Out-Degree*, representing outgoing connections from a vertex or the *Weighted In-Degree*, representing incoming connections to a vertex).

5.3.5 Key Findings Hierarchical Structuring

First, the *Descriptive Statistics* in section 5.3.1 revealed a number of interesting outlier observations. Plotting the *Weighted Out-Degree* against the *Weighted In-Degree* showed the structural importance of the China Education and Research Network (AS4538) and Tata Communications (formerly VSNL), (AS4755) for Vodafone and Aircel, respectively. The analysis of the *Clustering Coefficient* suggested the presence of a few

Autonomous Systems with strong mutual interconnections, an indicator of hierarchical structuring, as already the *k-core decomposition* in Chapter 4 showed. An analysis of those Autonomous Systems with the largest *Eigenvector Centrality* exposed those Autonomous Systems with the strongest influence on the mobile broadband operator networks. The estimated coefficients of the first stage (covering the econometrics models in Model 1 and Model 2) robustly indicate the relationships between *Clustering Coefficient* and *Eigenvector Centrality* and connectivity (given by the *Weighted Out-Degree* in Model 1 and the *Weighted In-Degree* in Model 2). These relationships, hence, strongly indicate the hierarchical structuring of the upstream Internet market.

5.4 Affordability and Quality of Service

Lack of affordability of mobile broadband Internet is considered a key issue underlying the low adoption rates in developing countries. Understanding and addressing the drivers of low affordability (e.g. for mobile broadband), and high relative prices, is of crucial relevance for helping in bridging existing digital divides and leveraging socio-economic opportunities in these countries (see sections 2.2.2 and 2.2.4). Hence in this section, we will consider whether the evidence we discussed in the previous section 5.3, might have an impact on the affordability of the mobile broadband price plans (defined as price per Megabyte) and on the price plan's Quality of Service (QoS). The following model, focussing on these relationships, is used to explore the last two *Working Hypothesis* as abducted in the Literature Review (section 2.5):

WH2: 'Tamil Nadu mobile broadband operators that show signs of a hierarchical upstream Internet market structure offer less affordable mobile broadband price plans to an end-user'.

WH3: 'Those Tamil Nadu mobile broadband operators that show signs of a hierarchical upstream Internet market structure provide a lower quality of service to an end-user'.

We start again by providing a Descriptive Analysis of the observations on price plans that we obtained from GSMOutlook (2015a, 2015b, 2015c). This is followed by the econometric Model 3, which will utilise the coefficients obtained in the previous estimation stage from Model 1 (dependent variable *Weighted Out-Degree*). In particular, the estimated coefficients, $\widehat{lclus}_{\text{Model 1}}$ and $\widehat{leige}_{\text{Model 1}}$ from the covariates for $\text{Ln}(\text{Clustering Coefficient})$ and $\text{Ln}(\text{Eigenvector Centrality})$ are used as proxies for the

level of hierarchical network structuring. These proxies are then related to the affordability.

With the *Clustering Coefficient*, we refer to an established tradition in the Computer Science literature (Vázquez, Pastor-Satorras and Vespignani, 2002, p.5) that identifies the estimated negative slope of the log-log relations between *Clustering Coefficient* and connectivity, as a marker for the level of hierarchical structuring in the analysed networks. Similarly, Choi, Barnett and Chon (2006, p.87-89) consider that the *Eigenvector Centrality* metric indicates hierarchical structuring, since this metric captured the vertex influence, obtained through stronger connections to other influential vertices (see also sections 2.3.3 and 3.4.2). However, in the following Models 3.2 and Model 4, we extend their work by splitting total connectivity into incoming and outgoing connectivity, as they clearly capture very different aspects of the original *traceroutes*' network connections. Hence, the incoming and outgoing connectivity metrics may play completely different roles in terms of the level of hierarchical network structuring, when considered in their relation with the *Eigenvector Centrality*. The following analysis will focus on these critical differences.

Our *Working Hypothesis* is that the price per Megabyte, as calculated upon GSMOutlook (2015a, 2015b, 2015c) data, is affected by the operator networks' levels of hierarchical structuring. In the following, these levels will be captured by the estimated coefficients representing the slopes of the log-log relations between *Clustering Coefficient* and connectivity, $\widehat{lclus}_{Model\ 1}$, and of the log-log relations between *Eigenvector Centrality* and connectivity, $\widehat{leige}_{Model\ 1}$. These relations were estimated in the first stage and are used as key proxies for capturing the levels of hierarchical structuring. In detail, Model 3 is using the coefficients estimated in Model 1 (based on *Weighted Out-Degree*), $\widehat{lclus}_{Model\ 1}$ and $\widehat{leige}_{Model\ 1}$. These are considered as proxies indicating the levels of hierarchical structuring. Model 4 is also using as proxy variables, the coefficients estimated in Model 2 (based on *Weighted In-Degree*), $\widehat{lclus}_{Model\ 2}$ and $\widehat{leige}_{Model\ 2}$. Similarly, these are considered as proxies indicating the levels of hierarchical structuring.

Lastly, this section concludes by focussing on the correlation between the level of hierarchical structuring, proxied by the estimated variables in Model 1 ($\widehat{lclus}_{Model\ 1}$ and $\widehat{leige}_{Model\ 1}$) and Model 2 ($\widehat{lclus}_{Model\ 2}$ and $\widehat{leige}_{Model\ 2}$), and the mobile broadband operator networks' Quality of Service (QoS). This metric is derived from the Telecom

Regulatory Authority of India. The correlation covers the *Working Hypotheses* WH3 stated above.

5.4.1 Descriptive Statistics

Given the available secondary data, the price plans of the three Tamil Nadu mobile broadband operators (Aircel, Bharti Airtel and Vodafone) were chosen as our dependent variable and unit of analysis. These data on price plans were collected to match the geo-location and dates of our data collection as described in the Methodology (section 3.3). A price plan, expressed in Indian Rupees (INR), includes information on:

- its duration (counted in days),
- the maximum data allowance (measured in Megabytes),
- a maximum connection speed (represented in 2G or 3G mobile technology) and
- the inclusion of added services (such as unlimited usage of Facebook or WhatsApp).

Table 5-21 below depicts the structure and examples of a typical price plan observation.

Price plan examples			
Operator	Price in INR	Validity in days	Price Plan description, examples given by GSMOutlook (2015a, 2015b, 2015c)
1	25	5	200MB 2G data
2	254	28	1 GB 3G + 150 MB Facebook + 200 MB Whatsapp
3	148	28	1GB 2G free mobile internet
Key GB = Gigabyte (=1024 MB). INR = Indian Rupees. MB = Megabyte (=1024 Byte). Operator (mobile broadband operator): 1 = Aircel, 2 = Bharti Airtel, 3 = Vodafone.			

Table 5-21: Price plan examples.

The Descriptive Statistics of the price plan observations in Table 5-22 below showed that the mean price per Megabyte of Tamil Nadu mobile broadband price plans (based on the three mobile broadband operators) was equal to INR 0.190523.

Descriptive Statistics, - Model 3 and Model 4					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
datainmb	46	2745.109	3669.781	25	15000
vin	46	21.19565	16.38918	1	90
price	46	381.8043	481.3748	8	2251
pricepermb	46	.190523	.0892192	.0509	.5183333
<p>Key</p> <p>datainmb = Data of price plan in Megabyte.</p> <p>price = Price in INR as stated in price plan.</p> <p>pricepermb = Price in INR as stated in price plan / Data of price plan in MB (Megabyte).</p> <p>vin = Validity of price plan in days.</p>					

Table 5-22: Descriptive statistics Model 3 and Model 4.

The following Table 5-23 - Table 5-25 illustrate the *Descriptive Statistics* for our three mobile broadband operators. Bharti Airtel provided, with a mean of 27.4 days, the highest duration in days for their price plans, followed by Aircel with 20.5 days and lastly Vodafone with 14 days. Moreover, Bharti Airtel showed a mean price plan price of INR 566.73, followed by Aircel with INR 402.6, and lastly Vodafone with INR 91.82, indicating their end-user focus. This indication is very interesting since Fennell et al. (2016) find, based on a survey in rural Tamil Nadu districts, that 39% of respondents were willing to spend between INR 27 to INR 400 per month (for 40MB – 750MB), while 21% were willing to spend between INR 100 to INR 225 per month (for a 1GB data allowance).

Aircel Descriptive Statistics - Price Plan					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
datainmb	20	4250	4389.251	100	15000
vin	20	20.5	11.82549	1	28
price	20	402.6	373.7165	25	1397
pricepermb	20	.1297294	.0591533	.0509	.2866667
<p>Key</p> <p>datainmb = Data of price plan in Megabyte.</p> <p>price = Price in INR as stated in price plan.</p> <p>pricepermb = Price in INR as stated in price plan / Data of price plan in MB (Megabyte).</p> <p>vin = Validity of price plan in days.</p>					

Table 5-23: Aircel descriptive statistics Model 3 and Model 4.

Bharti Airtel Descriptive Statistics - Price Plan					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
datainmb	15	2338.333	3121.482	25	12000
vin	15	27.4	22.57938	1	90
price	15	566.7333	669.7353	8	2251
Pricepermb	15	.2599468	.0853216	.1804	.5183333
<p>Key</p> <p>datainmb = Data of price plan in MB (Megabyte).</p> <p>price = Price in INR as stated in price plan.</p> <p>pricepermb = Price in INR as stated in price plan / Data of price plan in MB (Megabyte).</p> <p>vin = Validity of price plan in days.</p>					

Table 5-24: Bharti Airtel descriptive statistics Model 3 and Model 4.

Vodafone Descriptive Statistics - Price Plan					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
datainmb	11	563.6364	587.2993	25	2000
vin	11	14	10.65833	1	28
price	11	91.81818	68.97944	8	199
pricepermb	11	.2063881	.0638925	.0995	.32
<p>Key</p> <p>datainmb = Data of price plan in MB (Megabyte).</p> <p>price = Price in INR as stated in price plan.</p> <p>pricepermb = Price in INR as stated in price plan / Data of price plan in MB (Megabyte).</p> <p>vin = Validity of price plan in days.</p>					

Table 5-25: Vodafone descriptive statistics Model 3 and Model 4.

Looking at the price plots by provider in Figure 5-26 indicates that each operator showed an approximately normal distribution. Furthermore, the price plans covered a higher mean of data in Megabyte, see Figure 5-27, while Vodafone shows a higher density of shorter ranging price plans (measured in validity in days) as depicted in Figure 5-28.

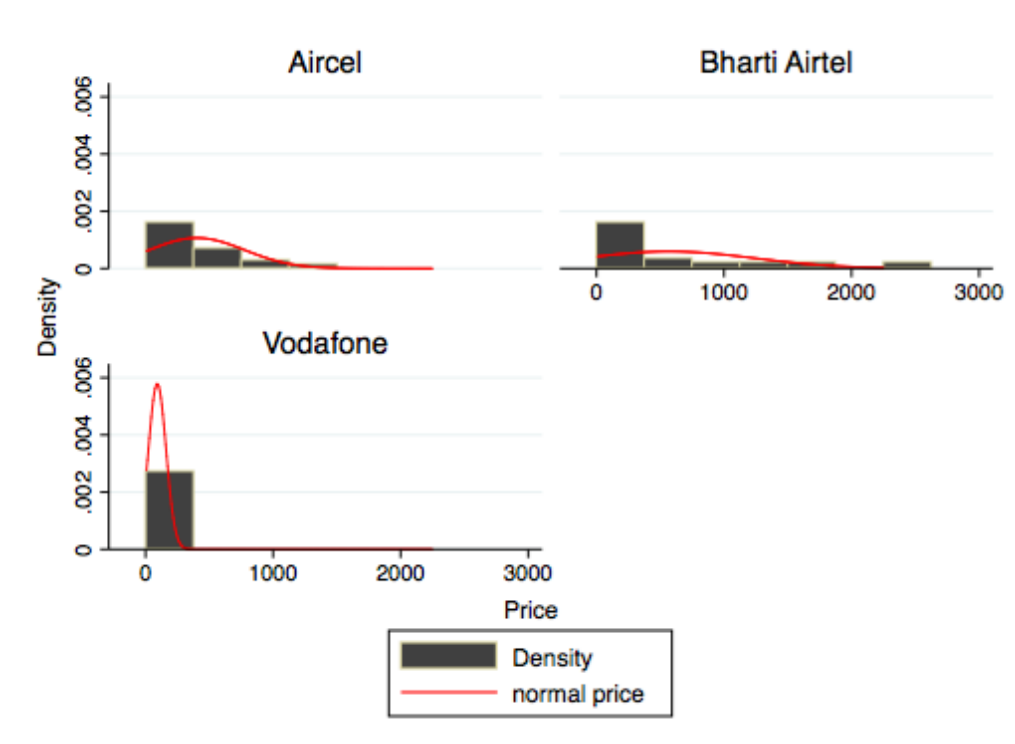


Figure 5-26: Price distribution by mobile broadband operator.

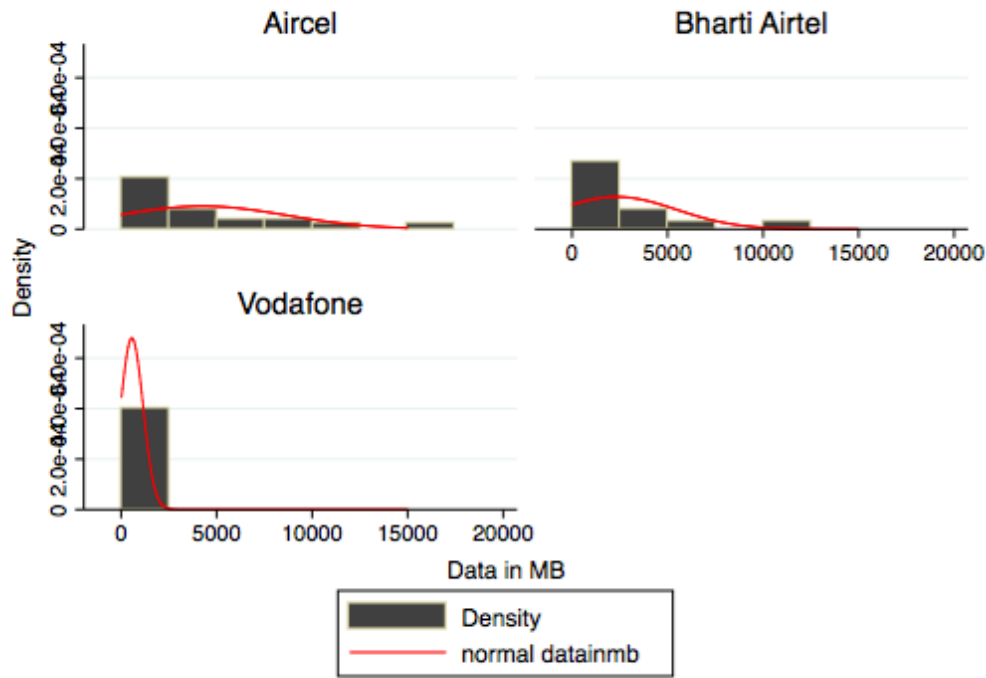


Figure 5-27: Data in Megabyte distribution by mobile broadband operator.

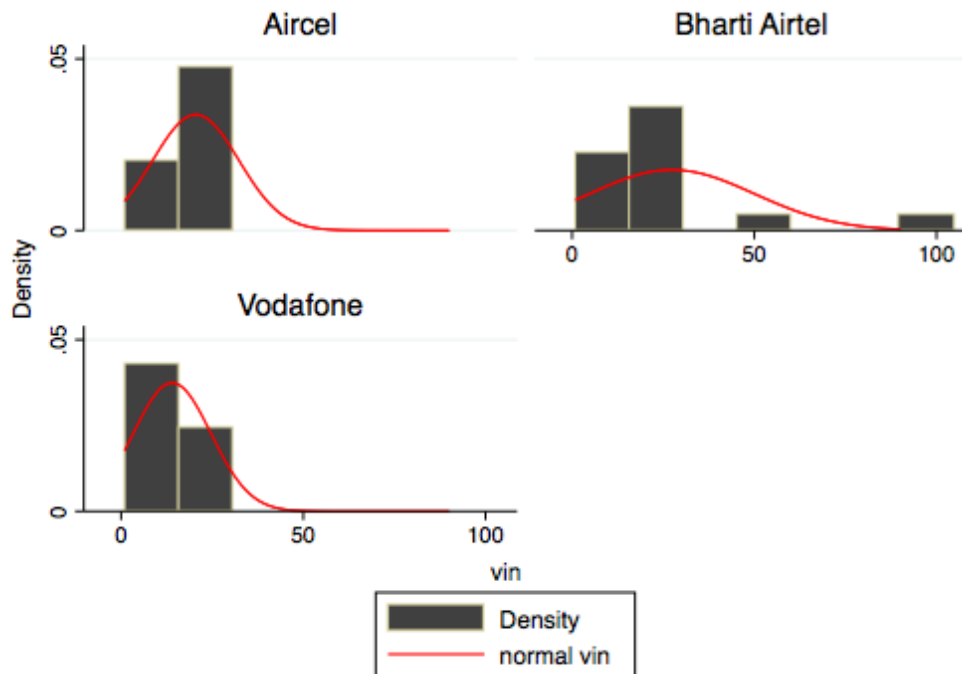


Figure 5-28: Validity in days distribution by mobile broadband operator.

Figure 5-29 below plots the price against data in Megabyte. We can see that the price distribution for Bharti Airtel, in comparison to both Aircel and Vodafone, shows a steeper value line. Interestingly, Vodafone offered no price plans with a large data allowance. Moreover, while Bharti Airtel showed a large service gap between 6GB of data and 12GB of data, Aircel seemed to provide the overall most-balanced price plan portfolio. Both Bharti Airtel and Aircel offered price plans, where the same amount of data in Megabytes was offered at different prices. Looking at the data as well as the detailed price plans revealed that their differences were given by the validity in days or added services such as unlimited access to services such as Facebook or WhatsApp.

Moreover, Bharti Airtel showed the highest mean in price per Megabyte (*pricepermb*), being price per Data in Megabyte for each operator-based price plan observations. Moreover, as shown in Figure 5-30, there were a number of Bharti Airtel observations with a large *pricepermb*. Nevertheless, Figure 5-30 on the next page reports the approximately normal distribution for price per Megabyte (*pricepermb*).

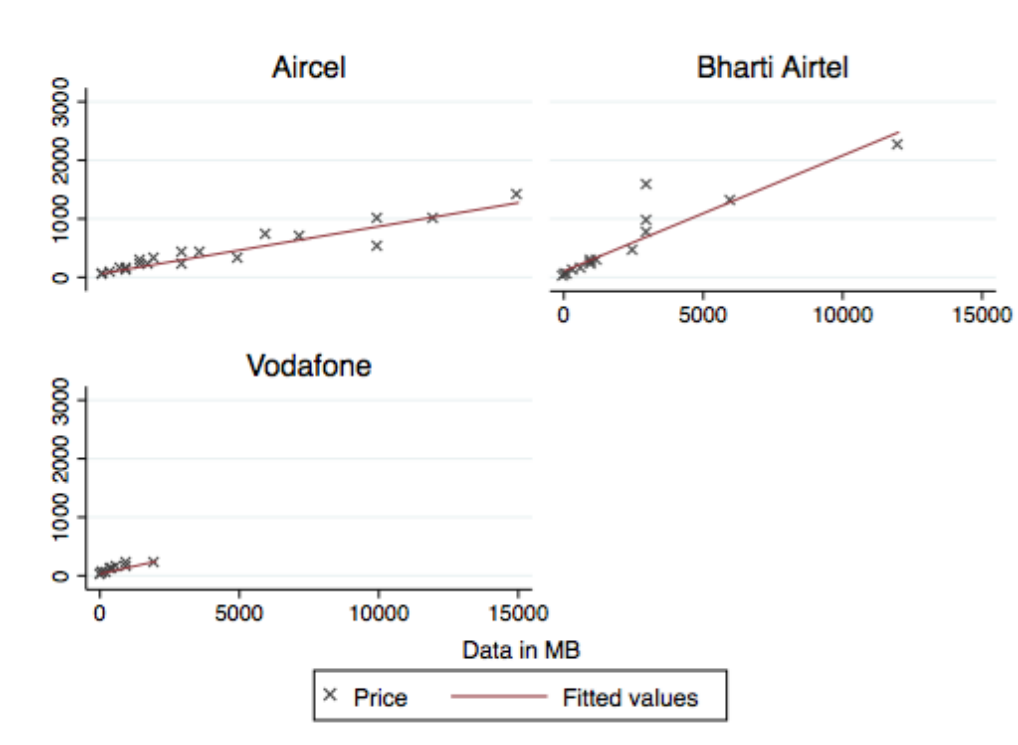


Figure 5-29: Two-way scatter plots data in Megabyte against price per mobile broadband operator with linear fit line.

Table 5-26 below reports, for each mobile broadband operator, the variables derived from the estimated coefficients from Model 1 and Model 2 (see sections 5.3.3 and 5.3.4 above), used as proxies for the operators' levels of hierarchical structuring.

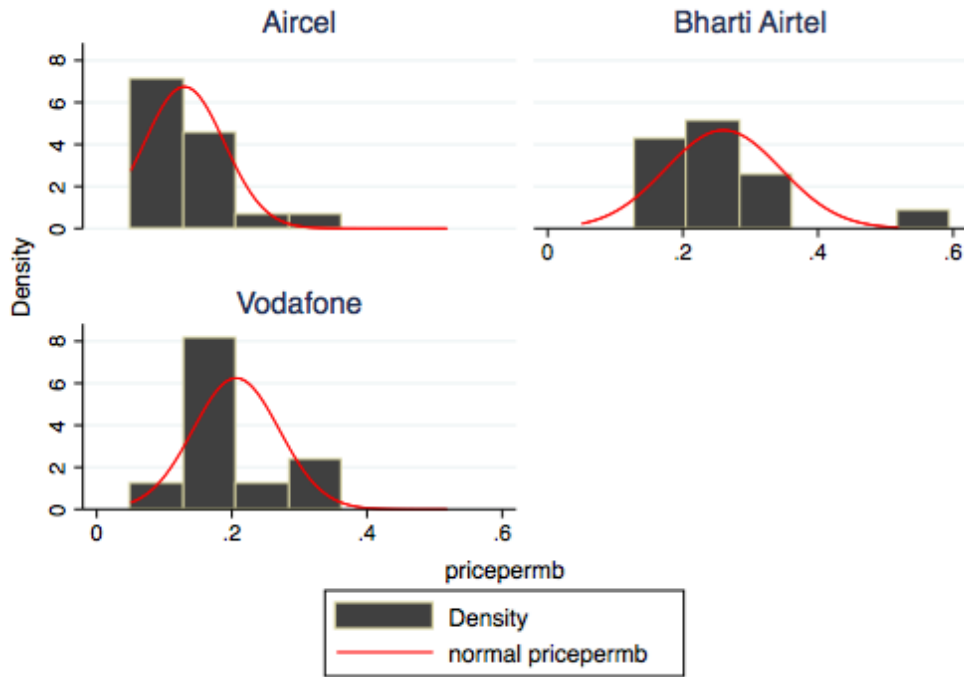


Figure 5-30: Price per Megabyte distribution by mobile broadband operator.

Categorical covariates derived from Model 1 and Model 2		
Categorical covariates derived from Model 1 and Model 2	Proxy estimated in Model 1 (<i>Weighted Out-Degree</i> based) per operator	Proxy estimated in Model 2 (<i>Weighted In-Degree</i> based) per operator
\widehat{lclus}_{Airtel}	-1.182964	-1.104983
\widehat{leige}_{Airtel}	.3212572	.334886
$\widehat{lclus}_{Bharti Airtel}$	-.9986846	-1.032788
$\widehat{leige}_{Bharti Airtel}$.277029	.2765935
$\widehat{lclus}_{Vodafone}$	-.8647558	-.7859873
$\widehat{leige}_{Vodafone}$.5424204	.5514546
<p>Key</p> <p>$\widehat{lclus}_{Operator}$: proxy, representing the estimated coefficient in Model 1 and 2 (Clustering Coefficient on Weighted Out-Degree), per operator.</p> <p>$\widehat{leige}_{Operator}$: proxy, representing the estimated coefficient in Model 1 and 2 (Eigenvector Centrality on Weighted Out-Degree), per operator.</p>		

Table 5-26: Categorical covariates derived from Model 1 and Model 2, used as proxies for hierarchical structuring per mobile broadband operator.

In order to capture the effect of level of hierarchical network structuring of the mobile broadband operators on their price per Megabyte (*pricepermb*), the proxies (estimated in the first stage of the two-stage econometric estimation) were used in this second stage as the covariates for the Model 3.1, Model 3.2 and Model 4 described below. In more detail, the following Model 3.1 and Model 3.2 include the proxies estimated in Model 1 (covering all three Tamil Nadu mobile broadband operators) above.

Model 3.1:

$$\ln(\text{pricepermb}_{\text{All operators}}) = \beta_0 + \beta_1(\widehat{lclus}_{\text{Model 1 (per operator)}}) + \varepsilon$$

Model 3.2:

$$\ln(\text{pricepermb}_{All\ operators}) = \beta_0 + \beta_1(\widehat{leige_{Model\ 1\ (per\ operator)}}) + \varepsilon$$

whereas, Model 4 includes those proxies estimated in Model 2 (again covering all three mobile broadband operators):

Model 4:

$$\ln(\text{pricepermb}_{All\ operators}) = \beta_0 + \beta_1(\widehat{lclus_{Model\ 2\ (per\ operator)}}) + \beta_2(\widehat{leige_{Model\ 2\ (per\ operator)}}) + \beta_3(\text{vin}_{All\ operators}) + \varepsilon$$

Hence, Model 3.1 captures the impact on price per Megabyte of the different level of hierarchical structuring of the three mobile broadband operators as proxied by the estimated coefficients of the relevant log-log model estimated in Model 1 in the first stage. Similarly, Model 3.2 captures the impact on price per Megabyte of the different level of hierarchical structuring of the three mobile broadband operators as proxied by the estimated coefficients of the relevant log-log model estimated in Model 2, also in the first stage.

Model 4 captures the impact on price per Megabyte of the different level of hierarchical structuring (*Clustering Coefficient* and *Eigenvector Centrality*) of the three mobile broadband operators as proxied by the estimated coefficients of the relevant log-log model estimated in Model 1 and Model 2 in the first stage, including the validity in days of a price plan as control variable.

As discussed above, $\widehat{lclus_{Model\ 1}}$ and $\widehat{leige_{Model\ 1}}$ represent the estimates obtained during the first estimation stage in Model 1 (see Table 5-26 above). Similarly, $\widehat{lclus_{Model\ 2}}$ and $\widehat{leige_{Model\ 2}}$ represent the estimates obtained during the first estimation stage in Model 2. In the following section we continue with the specification and the estimation of the Model 3.1 and Model 3.2 followed by Model 4 in section 5.4.3.

5.4.2 Model 3

Due to a perfect correlation and multicollinearity between the two obtained proxies, as discussed above, Model 3 is further split into Model 3.1 and Model 3.2. Model 3.1 considers the $\widehat{lclus}_{\text{Model 1}}$ as covariate of interest, whereas Model 3.2 focusses on the $\widehat{leige}_{\text{Model 1}}$.

Model 3.1

The dependent variable, $\ln(\text{pricepermb}_{\text{per operator}})$ is continuous, whereas $\widehat{lclus}_{\text{Model 1}}$, derived as discussed above, from the estimated parameters of Model 1, is the categorical proxy of interest. This covariate expresses the values of the coefficient of the log-log relation between the *Clustering Coefficient* and the *Weighted Out-Degree* connectivity, as estimated in the stage one (Model 1), separately for each one of our mobile broadband operators (Aircel, Bharti Airtel and Vodafone).

Here, Model 3.1 is estimated using an OLS log-log specification based on 46 price plan observations (see Table 5-27 below). The overall model is significant at 90% level (p-value of 0.0273 for $F(1, 44) = 5.21$). The model explains 8.56% of the total variation of the $\ln(\text{pricepermb}_{\text{per operator}})$ with a low Adjusted R-Squared equals to 0.0856.

The t-test showed that the categorical covariate, $\widehat{lclus}_{\text{Model 1}}$, capturing the level of hierarchical structuring of a mobile broadband operator network, estimated in the first stage of the process (Model 1), is statistically significant at 90%. Moreover, the coefficient for this covariate indicates that $\widehat{lclus}_{\text{Model 1}}$ had a positive effect on the $\ln(\text{pricepermb}_{\text{per operator}})$.

Model 3.1 Multiple Log-Log regression – Price plan Ordinary Least Squares (OLS)						
				Number of obs = 46 F(1, 44) = 5.21 Prob > F = 0.0273 R-squared = 0.1059 Adj R-squared = 0.0856 Root MSE = .46763		
lnpricepermb	Coef.	Std.Error	t	p > t	[95% Conf. Interval]	
$\widehat{lclus}_{\text{Model 1}}$	1.241779*	.543921 5	2.28	0.027	.1455769	2.33798
_cons	-.5195505	.551058 3	-0.94	0.351	-1.630135	.5910345
Key * p < .05; ** p < .01; *** p < .001 $\widehat{lclus}_{\text{Model 1}}$: proxy, representing the estimated coefficient in Model 1 (Clustering Coefficient on Weighted Out-Degree). $\ln pricepermb$: Ln-transformed price per Megabyte, $\ln(pricepermb_{\text{per operator}})$.						

Table 5-27: OLS log-log regression Model 3.1.

From our estimates in Table 5-27_{above}, we see that a 1% increase in this proxy, representing an **increase** (blue arrow for \widehat{lclus} in Figure 5-31 below) in the level of hierarchical structuring (flattening of the steepness of the red-dotted negative slope in Figure 5-31 below, captured by the log-log regression between *Weighted Out-Degree* and *Clustering Coefficient* in Model 1) **increases** the price per Megabyte by 1.24%. This increase in price per Megabyte is interpreted as a reduction in affordability.

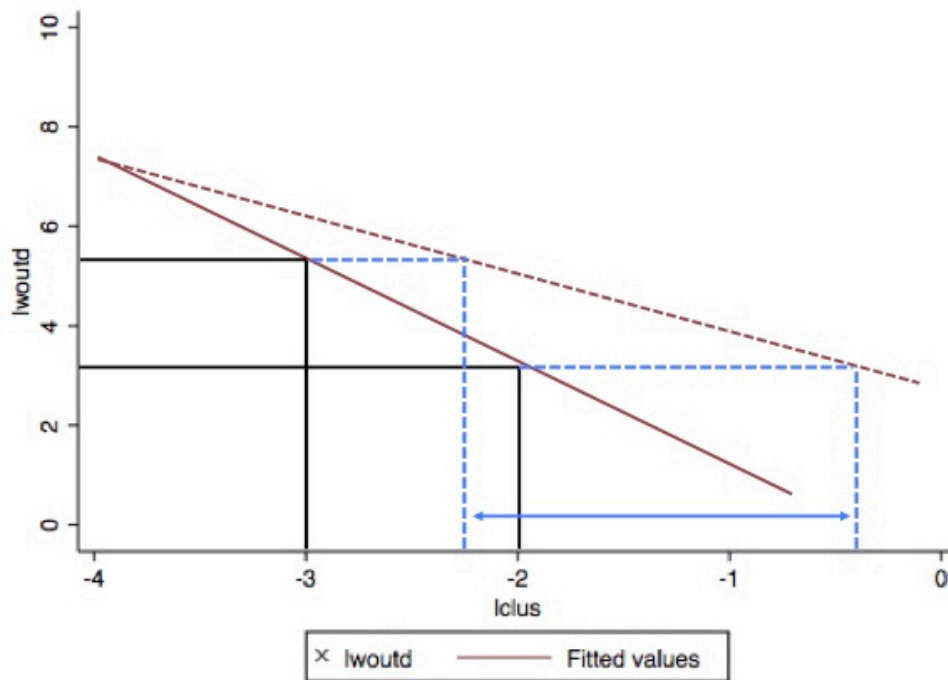


Figure 5-31: Increase in the *lclust_Model1_hat* proxy.

The Breusch-Pagan / Cook-Weisberg test for the estimated Model 3.1, showed a $\chi^2(1) = 1.43$ with a p-value of 0.2326 at normal significance levels. The Ramsey RESET test revealed, for $F(1, 43) = 29.61$, a p-value of 0.000. However, having carried additional estimations for more functional form specifications including higher power of the explanatory variables up to the fourth order (not reported here), we saw no improvement in this test, hence we again retained the original OLS log-log specification.

The estimated Variance Inflation Factor (VIF) was below the thresholds that would indicate the presence of multicollinearity. The plot of the predicted residual distributions, see Figure 5-32 below, shows an approximately normal distribution ranging closely around the mean of zero.

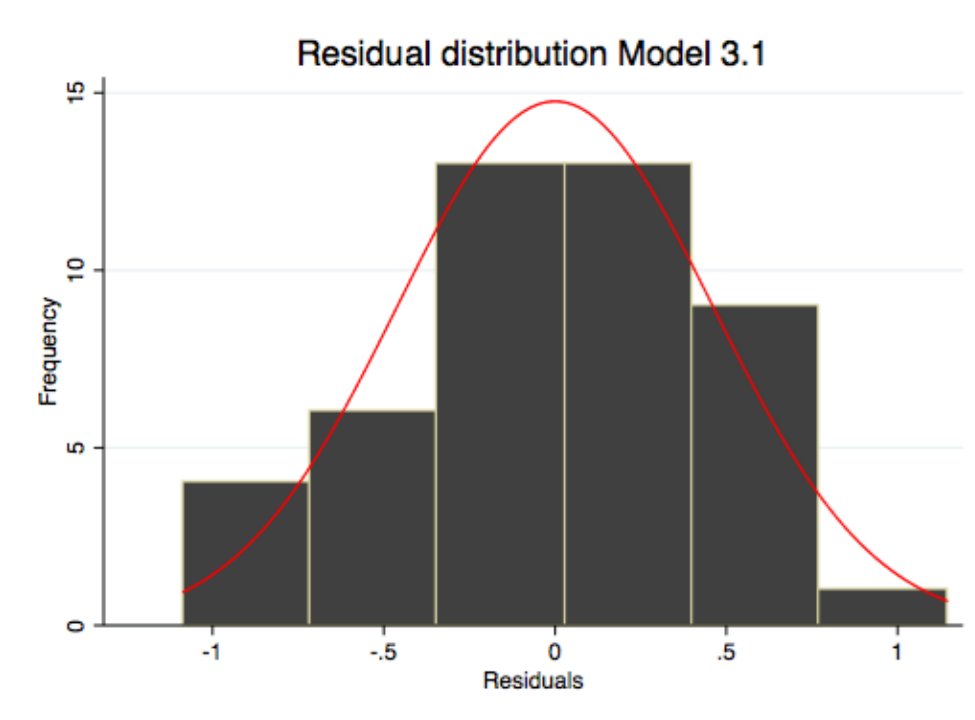


Figure 5-32: Predicted error term residual distribution Model 3.1.

Model 3.2

Here, the dependent variable, $\ln(\text{priceperm}_{\text{per operator}})$ is continuous, whereas $\widehat{\text{leige}}_{\text{Model 1}}$, derived covering all mobile broadband operators as discussed above, from the estimated parameters of Model 1, is the categorical covariate of interest. This covariate takes the values expressing the estimated coefficient of the log-log effect of the *Eigenvector Centrality*, on the *Weighted Out-Degree* connectivity, per operator.

Here, Model 3.2 is again, like Model 3.1 above, estimated using an OLS log-log specification based on 46 price plan observations (see Table 5-27 below). The overall model is significant at 90% level (p-value of 0.0327 for $F(1, 44) = 4.86$). The model explains 7.90% of the total variation of the $\ln(\text{priceperm}_{\text{per operator}})$ with a low Adjusted R-Squared equals to 0.0790.

The t-test showed that the categorical covariate, $\widehat{\text{leige}}_{\text{Model 1}}$, capturing the level of the hierarchical structuring of a mobile broadband operator network, estimated in the first stage of the process (Model 1), is statistically significant at 90%. Moreover, the coefficient for this covariate indicates that $\widehat{\text{leige}}_{\text{Model 1}}$ had a positive effect on the $\ln(\text{priceperm}_{\text{per operator}})$.

Model 3.2 Multiple Log-Log regression – Price plan Ordinary Least Squares (OLS)						
				Number of obs = 46 F(1, 44) = 4.86 Prob > F = 0.0327 R-squared = 0.0995 Adj R-squared = 0.0790 Root MSE = .4693		
lpriceperm b	Coef.	Std.Erro r	t	p > t	[95% Conf. Interval]	
$\widehat{leige}_{\text{Model 1}}$	1.393724*	.632058 8	2.21	0.033	.1198933	2.667555
_cons	-2.271297***	.238619 2	-9.52	0.000	-2.752202	-1.790391
Key * p < .05; ** p < .01; *** p < .001 $\widehat{leige}_{\text{Model 1}}$: proxy, representing the estimated coefficient in Model 1 (Eigenvector Centrality on Weighted Out-Degree). <i>lpriceperm</i> b: Ln-transformed price per Megabyte, $\ln(\text{priceperm}b_{\text{per operator}})$.						

Table 5-28: OLS log-log regression Model 3.2.

We interpret an increase in the slope of the log-log relation between the *Eigenvector Centrality* and *Weighted Out-Degree* in Model 1 as an increase in the level of hierarchical structuring since a steeper slope implies that more influential vertices (higher *Eigenvector Centrality* value) transport data packets, in one step, to more adjacent vertices, given the larger number of outgoing connections. A steeper slope in this interpretation implies that for a given value of *Eigenvector Centrality* there is a higher value of outgoing connectivity. Hence, these vertices can be seen as very central connectivity bottleneck that allow other vertices to spread many (one step) outgoing connections needed to reach the Internet Periphery.

Hence, from our estimates in Table 5-28_{above}, we see that a 1% increase in this proxy, representing an **increase** in the level of hierarchical structuring (sharpening of the steepness of the positive slope), captured by the log-log regression between *Weighted*

Out-Degree and *Eigenvector Centrality* in Model 1) **increases** the price per Megabyte by 1.39%, hence a reduction affordability.

The Breusch-Pagan / Cook-Weisberg test for the estimated Model 3.2, showed a $\chi^2(1) = 1.52$ with a p-value of 0.2171 at normal significance levels. The Ramsey RESET test revealed, $F(1, 43) = 30.13$, a p-value of 0.000. However, having carried additional estimations for more functional form specifications including higher power of the explanatory variables up to the fourth order (not reported here), we saw no improvement in this test, hence we again retained the original OLS log-log specification. The plot of the predicted residual distributions, see Figure 5-33 below, shows an approximately normal distribution ranging closely around the mean of zero.

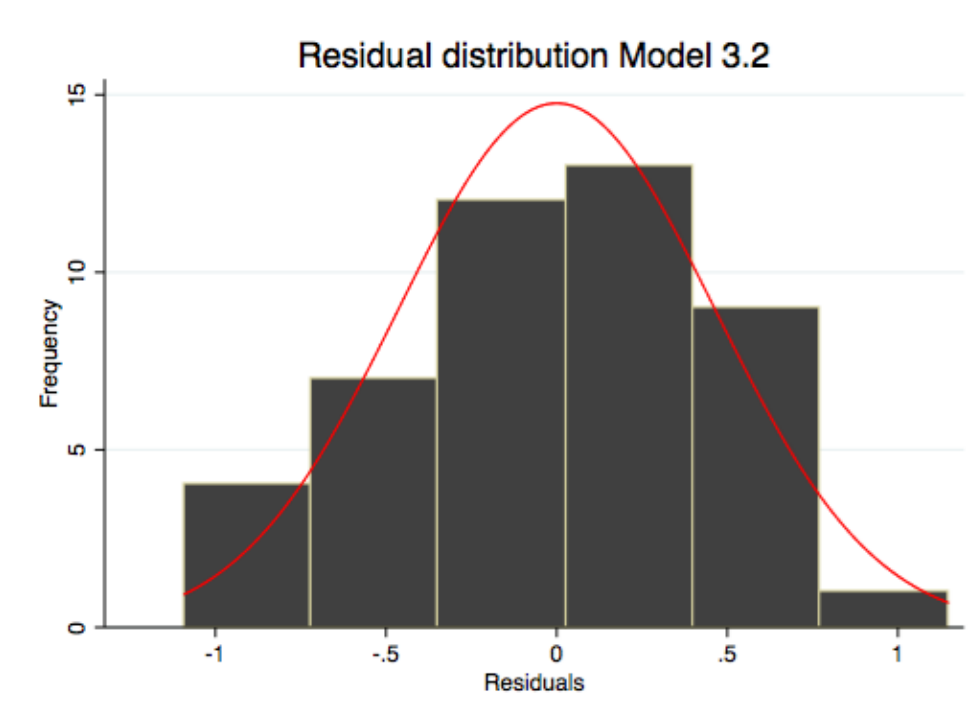


Figure 5-33: Predicted error term residual distribution Model 3.2.

Summary Model 3.1 and Model 3.2

The coefficient of the proxy covariate for the *Clustering Coefficient* ($\widehat{lclus}_{\text{Model 1}}$, derived from the estimated parameters of Model 1) in the estimated Model 3.1 above showed, that an **increase** in the level of hierarchical structuring (flattening of the steepness of the negative slope, captured by the log-log regression between *Weighted Out-Degree* and *Clustering Coefficient* in Model 1) results in an **increase** of price per Megabyte.

Interestingly, the coefficient of the proxy covariate for the *Eigenvector Centrality* ($\widehat{leige}_{\text{Model 1}}$, derived from the estimated parameters of Model 1) in the estimated Model 3.2 above showed, that an **increase** in the level of hierarchical structuring (sharpening of the steepness of the positive slope, captured by the log-log regression between *Weighted Out-Degree* and *Eigenvector Centrality* in Model 1) results in an **increase** of price per Megabyte.

This mismatch between the two proxy covariates (based on Model 1: level of hierarchical structuring on outgoing connectivity) is further explored in Model 4 (based on Model 2: level of hierarchical structuring on incoming connectivity) below.

5.4.3 Model 4

The dependent variable, $\ln(\text{pricepermb}_{\text{per operator}})$ is continuous, whereas $\widehat{lclus}_{\text{Model 2}}$ and $\widehat{leige}_{\text{Model 2}}$, derived covering all mobile broadband operators as discussed above, from the estimated parameters of Model 2, are the categorical covariates of interest. These covariates take the values expressing the estimated coefficient of the log-log effect of the *Clustering Coefficient* and the *Eigenvector Centrality*, on the *Weighted In-Degree* connectivity, per operator. Moreover, $\text{vin}_{\text{per operator}}$, the validity in days of a given price plan observation, is introduced as control variable.

Model 4 is estimated using an OLS log-log specification also based on 46 price plan observations (see Table 5-29 below). The overall model is significant (p-value of 0.0000 for $F(3, 42) = 12.93$). The model explains 44.30% of the total variation of the $\ln(\text{pricepermb}_{\text{per operator}})$ with an Adjusted R-Squared equals to 0.4430.

The t-test showed that the categorical covariates, $\widehat{lclus}_{\text{Model 2}}$ and $\widehat{leige}_{\text{Model 2}}$, capturing the effects of hierarchical structuring of a mobile broadband operator, estimated in the first stage of the process (Model 2), is statistically significant at 99%.

The coefficient for these covariates indicates that $\widehat{lclus}_{\text{Model 2}}$ had a positive effect on the $\ln(\text{pricepermb}_{\text{per operator}})$, whereas $\widehat{leige}_{\text{Model 2}}$ had a negative effect on the $\ln(\text{pricepermb}_{\text{per operator}})$. From our estimates in Table 5-29 above, we see that a 1% increase in the $\widehat{lclus}_{\text{Model 2}}$ proxy, representing an **increase** in the level of hierarchical structuring (flattening of the steepness of the negative slope, captured by the log-log regression between *Weighted In-Degree* and *Clustering Coefficient* in Model 2) **increases**

the price per Megabyte, $\ln(\text{priceperm}_{\text{per operator}})$, by 5.68%.

Model 4 Multiple Log-Log regression – Price plan Ordinary Least Squares (OLS)						
				Number of obs = 46 F(3, 42) = 12.93 Prob > F = 0.0000 R-squared = 0.4801 Adj R-squared = 0.4430 Root MSE = .36497		
$\ln(\text{priceperm}_{\text{b}})$	Coef.	Std. Error	t	p > t	[95% Conf. Interval]	
\widehat{lclus}	5.68335***	.912652	6.23	0.000	3.841544	7.525156
\widehat{leige}	-6.114105***	1.115993	-5.48	0.000	-8.366269	-3.86194
vin	-.0030792	.0034908	-0.88	0.383	-.010124	.0039656
_cons	6.25813***	1.308718	4.78	0.000	3.617051	8.899249
Key * p < .05; ** p < .01; *** p < .001 $\widehat{lclus}_{\text{Model 2}}$: proxy, representing the estimated coefficient in Model 2 (Clustering Coefficient on Weighted In-Degree), covering all three operators. $\widehat{leige}_{\text{Model 2}}$: proxy, representing the estimated coefficient in Model 2 (Eigenvector Centrality on Weighted In-Degree), covering all three operators. $\ln(\text{priceperm}_{\text{b}})$: Ln-transformed price per Megabyte, $\ln(\text{priceperm}_{\text{per operator}})$ vin: validity of price plan in days.						

Table 5-29: Multiple OLS log-log regression Model 4.

We interpret an increase in the slope of the log-log relation between the *Eigenvector Centrality* and *Weighted In-Degree* and in Model 2 as a decrease in the level of hierarchical structuring, because a steeper slope implies that influential vertices (higher *Eigenvector Centrality* value) receive data packets, from more adjacent vertices, because of a larger number of incoming connections. Hence, these vertices can be seen as

important hub-vertices that receive many incoming connections providing them with central access to the rest of the network.

Hence, we can see from the estimates in in Table 5-29 that a 1% increase in the $\widehat{leige}_{\text{Model 2}}$ proxy, representing an **decrease** in the level of hierarchical structuring (sharpening of the steepness of the positive), captured by the log-log regression between *Weighted In-Degree* and *Eigenvector Centrality* in Model 2) **decreases** the price per Megabyte, $\ln(\text{pricepermb}_{\text{per operator}})$, by 6.11%. Hence, both coefficients for our proxies from Model 2 show that a higher hierarchical structuring of a mobile broadband operator network (from a *Weighted In-Degree* perspective derived in Model 2) result in an **increase** of the price per Megabyte, hence in a **decreased** affordability.

The Breusch-Pagan / Cook-Weisberg test for the estimated Model 4, showed a $\chi^2(1) = 3.51$ with a p-value of 0.0611 at normal significance levels. The Ramsey RESET test revealed, for $F(3, 39) = 1.26$, a p-value of 0.3031. Hence, we again retained the original OLS log-log specification. The estimated Variance Inflation Factor (VIF) was again below the thresholds that would indicate the presence of multicollinearity. The plot of the predicted residual distributions, see Figure 5-34 below, shows an approximately normal distribution ranging closely around the mean of zero.

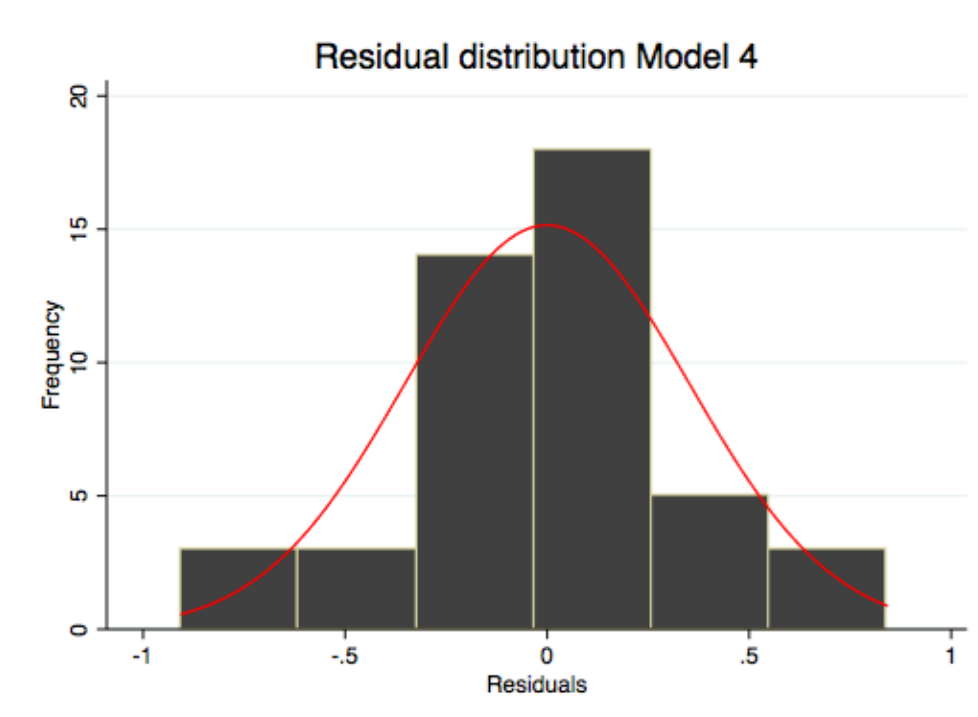


Figure 5-34: Predicted error term residual distribution Model 4.

Summarising, the coefficients of the two proxy covariates (from the estimated parameters of Model 2 (capturing the effect of the markers for the level of hierarchical structuring on *Weighted In-Degree* connectivity) in the estimated Model 4 consistently showed that a higher hierarchical structuring of a mobile broadband operator network (from a *Weighted In-Degree* perspective derived from Model 2) result in an **increase** of the price per Megabyte, hence in a **decreased** affordability.

5.4.4 Correlation with Quality of Service

Table 5-30 on the next page shows a correlation table between Quality of Service data, derived from the Telecom Regulatory Authority of India TRAI (2016a), for our three mobile broadband operators, and the categorical proxies estimated in Model 1 (see section 5.3.3) and Model 2 (see section 5.3.4) above.

Those proxies generated in Model 1 above are again referred to as $\widehat{lclus}_{Model\ 1}$ and $\widehat{leige}_{Model\ 1}$, incorporating all three mobile broadband operators. Similarly, those proxies generated in Model 2 above are again referred to as $\widehat{lclus}_{Model\ 2}$ and $\widehat{leige}_{Model\ 2}$, while also incorporating all three mobile broadband operators.

Here, the download speed (*speed_d*) was highly negatively correlated with the positive $\widehat{leige}_{Model\ 1}$ and $\widehat{leige}_{Model\ 2}$. Second, the *throughput*, being defined as Average Throughput for Packet data (in Kbps), was also negatively correlated with both $\widehat{leige}_{Model\ 1}$ and $\widehat{leige}_{Model\ 2}$ and $\widehat{lclus}_{Model\ 2}$. Third, the *pdp* as being the PDP Context activation success rate, >95% was positively correlated with $\widehat{lclus}_{Model\ 1}$. Lastly, the *drop_rate*, as being the package drop rate, <= 5%, was positively correlated with $\widehat{lclus}_{Model\ 1}$, $\widehat{leige}_{Model\ 1}$, $\widehat{lclus}_{Model\ 2}$ and $\widehat{leige}_{Model\ 2}$.

Quality of Service (QoS) Correlation Table				
	$\widehat{lclus}_{\text{Model 1}}$	$\widehat{leige}_{\text{Model 1}}$	$\widehat{lclus}_{\text{Model 2}}$	$\widehat{leige}_{\text{Model 2}}$
$\widehat{lclus}_{\text{Model 1}}$	1.0000			
$\widehat{leige}_{\text{Model 1}}$	0.7173	1.0000		
$\widehat{lclus}_{\text{Model 2}}$	0.9222*	0.9310*	1.0000	
$\widehat{leige}_{\text{Model 2}}$	0.9163*	0.9363*	0.9999*	1.0000
sprov	-0.0590	0.6532	0.3318	0.3457
speed_d	-0.4324	-0.9384**	-0.7475	-0.7573**
throughput	-0,6211	-0,9916**	-0.8759**	-0.8830**
latency	0.4410	-0.3090	0.0595	0.0447
pdp	0.8959*	0.3332	0.6544	0.6431
drop_rate	0.7608*	0.9989*	0.9526*	0.9570*
<p>Key</p> <p>* high positive correlation ($>.75$),</p> <p>** high negative correlation ($<-.75$)</p> <p>$\widehat{lclus}_{\text{Model 1}}$: proxy, representing the estimated coefficient in Model 1 (Clustering Coefficient on Weighted Out-Degree), covering all three operators.</p> <p>$\widehat{lclus}_{\text{Model 2}}$: proxy, representing the estimated coefficient in Model 2 (Clustering Coefficient on Weighted In-Degree), covering all three operators.</p> <p>$\widehat{leige}_{\text{Model 1}}$: proxy, representing the estimated coefficient in Model 1 (Eigenvector Centrality on Weighted Out-Degree), covering all three operators.</p> <p>$\widehat{leige}_{\text{Model 2}}$: proxy, representing the estimated coefficient in Model 2 (Eigenvector Centrality on Weighted In-Degree), covering all three operators.</p>				

Table 5-30: Quality of Service (QoS) correlation table.

5.4.5 Key Findings Affordability and Quality of Service

The coefficients of the estimated Model 3.1, Model 3.2 Model 4 indicate the relationships between the hierarchical structuring of the upstream Internet access market and the affordability (described in price per Megabyte). Model 3.1 showed that an increase in the hierarchical structuring (based on the effect of the *Clustering Coefficient on Weighted Out-Degree* connectivity) results in an increase of price per Megabyte of the price plans. Model 3.2 confirmed this insight for the higher hierarchical structuring (based on the

effect of the *Eigenvector Centrality* on *Weighted Out-Degree* connectivity). Model 4 also confirmed this insight for connectivity based on the *Weighted In-Degree*. Lastly, the correlation indicated that the hierarchical structuring markers are also correlated with a number of Quality of Service (QoS) metrics.

5.5 Key Findings Statistical Network Analysis

Based on Model 1 and Model 2 in the first stage, the analysis of the hierarchical structuring revealed that a stronger hierarchical structuring of the upstream Internet access market of our three Tamil Nadu mobile broadband operators results in a stronger connectivity (measured based on *Weighted Out-* and *In-Degree* connectivity). Moreover, Model 3.1, Model 3.2 and Model 4 in the second stage indicated that the estimated relationship between hierarchical structuring and connectivity from the first stage has an effect on the affordability of mobile broadband (as measured in price per Megabyte). Here, an increase in hierarchical structuring was indicated to increase the price per Megabyte. Similarly, a reduction in hierarchical structuring would decrease the price per Megabyte. Therefore, we assume that a more hierarchical upstream Internet access market results in less affordable mobile broadband price plans available for end-users in the Internet periphery.

5.6 Summary Statistical Network Analysis

The first part of this chapter provided a number of necessary theoretical background information and descriptive statistics, followed by the specification of our econometric models. Next, this chapter provided the statistical network analysis of the hierarchical network structuring in section 5.3. Here, we explained Model 1 (based on *Weighted Out-Degree* connectivity) and Model 2 (based on *Weighted In-Degree* connectivity) of our two-stage econometric estimation process. Next, this chapter covered the relation between the identified hierarchical network structuring and affordability in section 5.4. Here, we elaborated the Model 3 (3.1 and 3.2) and Model 4, again separated between incoming and outgoing connectivity. Additionally, we elaborated a correlation table between the structural indicators and Quality of Service. In the last part of this Chapter, we restated the key findings of our two-staged econometric estimation process. Our elaborated findings will be critically discussed with the existing relevant literature in the following Chapter 6, while Chapter 7 will state our identified key contributions to knowledge and its related limitations.

6 DISCUSSION

*‘I do not fear the Internet. I fear its understatement and unequal access.’
(Sigloch, 2017).*

The analysis presented in the previous chapters evidenced key features indicating a hierarchical nature of the three Tamil Nadu mobile broadband operator networks, Aircel, Bharti Airtel and Vodafone. These features also shed light on these operators’ upstream internetworking reliance on certain *Internet Service Providers*, some of whose Autonomous System relationships were previously unknown. This indicated the presence of a positive association between price per Megabyte and the degree of hierarchical structuring for these mobile broadband operator networks (measured through the two *Eigenvector Centrality* and *Clustering Coefficient* proxies). The following discussion aspires to support the dissertation’s main aim, by placing our findings into the relevant research context, to evidence our main contributions to knowledge while providing recommendations to policy and practice in Chapter 7.

The sections below revisit the *Working Hypotheses* that were derived from the Literature Review (section 2.5). We start, in section 6.1, to relate our case study’s key evidence, indicating the presence of hierarchical structuring in the three Tamil Nadu mobile broadband operator networks, to *Working Hypotheses* WH1, WH1.1 and WH1.2, discussed in section 2.5. Next, section 6.2 discusses the effects of the hierarchical structuring of the mobile broadband operator networks on the affordability of end-user price plans, as hypothesised in *Working Hypothesis* WH2. This is followed by a discussion of correlations with metrics capturing elements of Quality of Service addressing our *Working Hypothesis* WH3 in section 6.3. Lastly, in section 6.4, we discuss the implications to policy and practice before stating, in Chapter 7, our key contributions to knowledge, our case study limitations, ending with the conclusions and recommendations for future research endeavours.

6.1 Hierarchical Structuring

Based on our insights from the Literature Review in Chapter 2, we abducted the first Working Hypothesis WH1 by inferring that ‘The Tamil Nadu mobile broadband

operators' upstream Internet market structure displays features of a hierarchical ordering'. Next, we inferred that such hierarchical Internet market structuring could mean that 'The Tamil Nadu mobile broadband operators rely on an identified set of specific *Internet Service Providers* for their upstream connectivity' (WH1.1). Moreover, we hypothesised in WH1.2 that 'Studying the Tamil Nadu mobile broadband operators from an Internet-Periphery perspective indicates previously hidden upstream AS relationships'. Our evidence from Chapters 4 and 5 provides a multitude of findings related to these hypotheses featuring novel properties and a range of implications for policy and practice.

The following paragraphs report our key findings and discuss the evidence in light of the relevant literature and recent advancements. In detail, section 6.1.1 below looks at the apparent hierarchical structuring of the three Tamil Nadu mobile broadband operator networks. This is followed by a discussion of the upstream connectivity reliance upon certain Autonomous Systems and by the identification of previously hidden upstream Internet connectivity Autonomous System relationships. Finally, we summarise the key evidence of our case study.

6.1.1 Hierarchical Upstream Internet Access Market

The structure of internetworking amongst providers operating in the upstream Internet market is determined by the collection of their bilateral business relationships. These relationships detail the rules for the acceptance and re-routing of data traffic to reach destinations residing in the global Internet.

Among the Network metrics discussed in Chapter 4, we have seen that the *Clustering Coefficient* captures the existence of mutual interconnectivity among neighbours of an Autonomous System, i.e. the probability that any two of them are connected between themselves (Watts & Strogatz, 1998). Interestingly, we found that Vodafone's network displays a considerably higher average *Clustering Coefficient* than those of Aircel or Bharti Airtel. The highest individual *Clustering Coefficient* was displayed by GREE Inc., an Autonomous System belonging to the Aircel's network, followed by a number of other Autonomous Systems with slightly lower *Clustering Coefficient* values. Bharti Airtel's network has a set of different individual Autonomous Systems, which share the highest *Clustering Coefficients*, including the Academic Computer Center TASK at the Technical University of Gdansk in Poland and Verizon Business (UUnet). Lastly, the Autonomous Systems with the highest *Clustering Coefficient* in Vodafone's network is the U.S.

Federal Reserve Board, followed by other Autonomous Systems including Norlisk Telecom JSC in Russia and Bayanat Al-Oula in Saudi Arabia. Nevertheless, all three mobile broadband operators still incorporate only a low number of Autonomous Systems with high *Clustering Coefficients*. This means that most of the identified Autonomous Systems do not strongly interconnect with their neighbours, reinforcing the relevance of certain connectivity paths with frequently occurring relationships between a few of them. This indicates the presence of hierarchical structuring, as discussed in relation to the literature reviewed for the development of WH1 in Chapter 2.

As a consequence, our evidence, in accordance with WH1, shows that a few existing business partnerships between Autonomous Systems are frequently visited and are therefore very important in reaching global connectivity. A practical implication of this is that mobile broadband operator efforts to increase routing efficiencies might be hampered due to the existence of strong bilateral business relationships as revealed by the data discussed in Chapter 4.

Moreover, the evidence on the Network metrics discussed in Chapter 4 shows the relevance of using the *Eigenvector Centrality* metric for undertaking structural analysis when assessing the influence of certain Autonomous Systems within a network of interconnections generated by a mobile broadband operator network. This *Eigenvector Centrality* is calculated based on the concept that relationships to high-scoring Autonomous Systems contribute more to a given vertex influence than relationships to low-scoring ones.

The analysis in Chapter 4 also shows that the Autonomous Systems with the largest *Eigenvector Centralities* in the Aircel's network are Tata Communications (America) Inc., Level 3 Communications Inc., Cogent Communications, and NTT America Inc., most of which are large Tier-1 *Internet Service Providers*. The Autonomous Systems with the largest *Eigenvector Centrality* values in the Bharti Airtel's network are Bharti Airtel Ltd. Itself. This is followed by Level 3 Communications Inc., Breeze Network, Cogent Communications, NTT America Inc., the Amsterdam Internet Exchange (AIX), Hurricane Electric Inc. and Transtelecom. It is relevant to note that Bharti Airtel is the only mobile broadband operator that is strongly influenced by an Internet Exchange Point (IXP). Lastly, the Autonomous Systems with the largest *Eigenvector Centrality* values in the Vodafone network are Vodafone India Ltd., followed by Bharti Airtel Ltd., Telstra Global and Vodafone's Cable and Wireless Worldwide plc. subsidiary. The structural

importance of a few Tier-1 *Internet Service Providers* comes as no surprise given the high ranking they reach in the CAIDA (2016a) AS-Rank Database. These Autonomous Systems occupy influential internetworking positions for our three Tamil Nadu mobile broadband operator networks. Our findings point to the relevance of exploring upstream Internet structures using the *Eigenvector Centrality* as a proxy for hierarchical structuring. While Ruiz and Barnett (2015) utilise this metric, they fail to indicate its usefulness to study the hierarchical of the upstream Internet market. Unsurprisingly, the *k-core decomposition* discussed in Chapter 4 shows that most of these Autonomous Systems are also densely connected vertices in the three mobile broadband operator networks, while also discovering the presence of additional influential Autonomous Systems (discussed below). The similarity of results obtained through the analysis of *Eigenvector Centrality* and *k-core decomposition* provides additional robustness and value to our empirical findings. Moreover, our econometric models developed in stage one of Chapter 5 show that the relationship of *Clustering Coefficient* and *Eigenvector Centrality* to the upstream internetworking connectivity of Autonomous Systems (for both incoming and outgoing connections as measured by the *Weighted In- and Out-Degree* connectivity) indicates previously undiscovered features of the hierarchical structuring of the Tamil Nadu's mobile broadband operators' upstream markets. Our novel application of combining these two Network metrics to study these networks' hierarchical structuring in Model 1 and Model 2 adds to the existing body of knowledge (see Vázquez, Pastor-Satorras and Vespignani (2002), Dall'Asta et al. (2005) and Alvarez-Hamelin et al. (2008)). Our findings are in agreement with the work of Vázquez, Pastor-Satorras and Vespignani (2002, p.11), confirming the role of the *Clustering Coefficient* and scaling with a negative exponent of the connectivity as an indicator of the hierarchical organisation of the upstream Internet market structure. This shows that the many Autonomous Systems that connect to a few large International *Internet Service Providers* are not well connected amongst themselves. Moreover, given our econometric results obtained in Model 1 and Model 2 in Chapter 5, we show that the *Eigenvector Centrality* scales with a positive exponent in a log-log specification with connectivity (*Weighted In- and Out-Degree*), a key and novel empirical finding of our dissertation. Hence, we consider the relationship between *Eigenvector Centrality* and connectivity as a powerful additional indicator for understanding the hierarchical organisation of the upstream Internet market structure. It represents the market and bargaining powers of key influential Autonomous Systems for global connectivity. Moreover, our econometric

models add to the preliminary use of the *Clustering Coefficient* by Giovannetti and Sigloch (2015), who discover a hierarchical upstream Internet market structure of the B-Mobile mobile broadband operator in Bhutan. Adding to our analysis from Chapter 4, this evidence points towards the presence of hierarchical upstream Internet market structures for each one of the three Tamil Nadu mobile broadband operator networks. The hierarchical structuring of the upstream Internet market refers to situations where large Tier-1 *Internet Service Providers* benefit from economies of scale while securing superior market and bargaining powers, compared to smaller Tier-2 or Tier-3 *Internet Service Providers* (Frieden, 2001, p.362-368). As explained in section 5.3, our specified Model 1 indicates that an increase in the negative slope of the *Clustering Coefficient* results in a decrease of *Weighted Out-Degree* connectivity, as indicated by Vázquez, Pastor-Satorras and Vespignani (2002). However, an increase in the positive slope of the *Eigenvector Centrality* results in an increase of the *Weighted Out-Degree* connectivity (see section 5.3.3).

Moreover, we acknowledge that the relevance of the *Eigenvector Centrality* lies in its contribution towards the overall representation of the network structure due to the relevance it attributes to the total relationships in a given network (Tranos, 2013, p.93). Choi, Barnett and Chon (2006, p.87-89) already consider the *Eigenvector Centrality* as a metric that indicates hierarchical structuring. We agree with this view and consider the *Eigenvector Centrality* metric to be valuable since it considers certain relationships as more important, due to their frequency of occurrence. Hence, the *Eigenvector Centrality* does not attribute great value to relationships between Autonomous Systems that are not frequently visited. Such relationships could still be of structural importance (Borgatti and Li, 2009, p.10), e.g. bridging loosely-connected parts of the upstream operator networks. Those Autonomous Systems would clearly represent connectivity bottlenecks that should be carefully examined and understood by the network operational strategies. Given the statistical significances of the six models introduced in Chapter 5, to the best of our knowledge, we find a neglected and under-researched application of these metrics in both the Internet Economics and Computer Sciences.

6.1.2 Structural Bottlenecks

Structural bottlenecks of mobile broadband operator networks refer to the importance of certain *Internet Service Providers* (represented by their Autonomous Systems) for upstream internetworking and when providing connectivity to the global Internet. We can see how structural bottlenecks may cause traffic flow congestions, which would occur when an Autonomous System (or network node) receives more data traffic than it can cope with, similar to a motorway at rush hour. We refer to structural bottlenecks as being those Autonomous Systems that provide connectivity to the most densely connected core of large Tier-1 *Internet Service Providers* as well as among themselves (referring to Economides, 1995). In turn, these structural bottlenecks also interconnect traffic to the Internet periphery, as captured by the observed values of their *Eigenvector Centrality* in the analysed networks studied in Chapter 5.

The *Complex Network Analysis* performed in Chapter 4 provided the first clues about the existence of hierarchical structuring of the upstream Internet market for each Tamil Nadu mobile broadband operator. Meanwhile, the *Statistical Network Analysis* performed in Chapter 5 strengthened these initial evidences. By using the *k-core decomposition*, which focuses on properties such as network regions of increasing centrality and connectedness that represent network hierarchy (Alvarez-Hamelin et al. 2008, p.391), Chapter 4 also identified those Autonomous Systems that were distinctively and densely connected amongst each other. This represents key upstream Autonomous Systems for providing global connectivity to the Internet periphery. Moreover, these Autonomous Systems might have a strong market and downstream bargaining power over smaller (Tier-2 or Tier-3) Autonomous Systems, which appear in the networks generated by our three-studied mobile broadband operators.

The routing policies and properties of these key Autonomous Systems would potentially allow our mobile broadband operators to find beneficial upstream internetworking connectivity paths that employ specific Quality of Service properties, as Alvarez-Hamelin et al., (2005b; 2008) already indicate. Kang and Gligor (2014, p.10) argue that most BGP routings of Autonomous Systems would favour minimum-cost links over a uniform distribution of routes across several edges, covering a multitude of upstream Autonomous Systems. The *power-law degree distributions* identified in Chapter 4, where a few Autonomous Systems obtain the most connectivity (as measured in *Degrees*), support this claim by Kang and Gligor (2014). This evidence shows that our empirical findings

emerge from networks characterised by a few Autonomous Systems with high connectivity. Hence, our identified hierarchical structure of the three upstream mobile broadband operator networks relates to minimum-cost links, where routing is centralised at the network core, as identified through the *k-core decomposition* at Autonomous System granularity. Nevertheless, we note that this could be due to the randomisation process of the Portolan (2015) application. Real-world end-user, browsing patterns would provide stronger evidence regarding local connectivity, which would not necessarily target connections needing transit through the core Autonomous Systems of the Internet.

Unsurprisingly, the most densely connected Autonomous Systems (highest *k-core*) in the Aircel network are Tata Communications (America) Inc., Cogent Communications, NTT America Inc., Level 3 Communications, Telia Company AB, Cable and Wireless Worldwide plc and PCCW Global. The Bharti Airtel network shows Bharti Airtel Ltd., Cogent Communications, Level 3 Communications Inc., NTT America Inc. and Telia Company AB as most densely connecting Autonomous Systems. Lastly, the Autonomous Systems in the highest *k-cores* for the Vodafone network are Vodafone India Ltd., Cable and Wireless Worldwide Plc. (Vodafone subsidiary), the China Education and Research Network, Cogent Communications, Level 3 Communications Inc., NTT America Inc., PCCW Global and lastly AT&T Services Inc. Our three studied Tamil Nadu mobile broadband operators rely on those Autonomous Systems for their global internetworking connectivity purposes. Most of these Autonomous Systems are very highly ranked in the CAIDA (2016a) AS-Rank and most of them are identifiable by using the *Eigenvector Centrality* as described above.

Our evidence adds to Alvarez-Hamelin et al. (2005b; 2008), who relate the hierarchical structuring to the role that vertices play in terms of centrality and connectivity patterns, where connectivity relates to robustness (against faults and cyber-attacks) and routing (Quality of Service and efficiency). We show that a low number of Autonomous Systems (belonging to large *Internet Service Providers*) are extremely densely interconnected between themselves, while also providing crucial links to more peripheral and less well-connected Autonomous Systems residing in the Internet periphery. Therefore, our three Tamil Nadu mobile broadband operators seem reliant on the National (Tier-2) and International (Tier-1) *Internet Service Providers* identified above. The *k-core decomposition* to study Internet structures at AS granularity is introduced by Gaertler and Patrignani (2004), Alvarez-Hamelin et al. (2005b; 2008) and Dorogovtsev, Goltsev and

Mendes (2006). Additional examples include Fay et al. (2010), Alvarez-Hamelin, Beiró and Busch (2011) and Gregori, Lenzini and Orsini (2013). Interestingly, there is very little attempt to analyse the hierarchically-structured economic situations of such key Autonomous System interconnections using the *k-core decomposition* (except for CAIDA's regular application). One reason for this lack of research could be accounted for by the usual application of the *k-core decomposition* in biological settings to analyse and predict protein interactions (Alvarez-Hamelin et al. 2008). Only Garas et al. (2010) use the *k-core decomposition* to analyse situations of economic crisis. Hence, our work provides a valuable step in an emerging direction, while also uniquely indicating the key players for the upstream connectivity of our three operator networks.

To the best of our knowledge, no other research effort uses a methodological combination of *Complex Network Analysis*, *Graph Visualisation Analysis* (including the *k-core decomposition*) and *Statistical Network Analysis* to study the upstream Internet market structure based on active Internet periphery measurements at both IP and AS granularity, while linking the obtained results to study economic affordability in a lower-middle income country such as India.

This work introduces a strong and novel methodological approach to exploring the hierarchical structuring and structural bottlenecks in the upstream Internet market. However, the underlying economic relationships between the key Autonomous Systems remain hidden when only using the *k-core decomposition*, since this method does not account for connectivity directions of relationships (see also Alvarez-Hamelin et al. 2008, p.373). Our additional evidence, discussed below, aims to both detect the connectivity bottlenecks as well as considering the economic implications for these key relationships. This is achieved by fusing our primary *traceroute* data from the active Internet periphery measurements with the secondary CAIDA (2016b) AS-relationship dataset.

Given the evidence from Chapter 4 and Chapter 5, we do not reject Working Hypothesis WH1: 'The Tamil Nadu mobile broadband operators' upstream Internet market structure displays features of a hierarchical ordering'. Moreover, we also do not reject Working Hypothesis WH1.1: 'The Tamil Nadu mobile broadband operators rely on an identified set of specific Internet Service Providers for their upstream connectivity'.

6.1.3 Hidden Autonomous System Relationships

The joint consideration of the above findings (*k-core decomposition* at AS granularity)

with the secondary CAIDA (2016b) AS-Relationship dataset provided an initial picture of the economic nature of the upstream Internet market while indicating key structural bottleneck relationships. The following paragraphs describe and discuss these findings.

Firstly, most of Aircel's fused Autonomous System relationships are of a *provider-to-customer* nature (3.94% of the total relationships), followed by *peer-to-peer* relationships (3.33%), and lastly *customer-to-provider* ones (2.17%). These findings indicate that the Aircel upstream Internet market incorporates a considerable amount of paid transit connections (including paid *peer-to-peer* relationships). Plotting an edge-weighted graph visualisation (see Figure 4-21) into a two-dimensional Euclidean space exposes that the Aircel's upstream Internet market highlights a strong *peer-to-peer* relationship between Tata Communications (America) Inc. and the Tier-1 Autonomous System Level 3 Communications Inc. Moreover, our exploration finds two sets of frequently used *customer-to-provider* connections. The first between Tata Communications (America) Inc. and the Tier-1 Cogent Communications, and the second between Tata Communications (America) Inc. and NTT America Inc., another Tier-1 *Internet Service Provider*. This indicates that Aircel faces substantial upstream connectivity reliance on Tata Communications (America) Inc. This reliance could potentially force Aircel to cover the downstream costs that their upstream partner pays when forwarding data packets through the two upstream *Internet Service Providers* for global connectivity purposes, indicating the bottleneck position of Tata Communications (America) Inc.

Secondly, most of the fused Autonomous System relationships are of *peer-to-peer* nature (2.51% of the total relationships), followed by *provider-to-customer* (2.30%), and lastly by *customer-to-provider* (0.42%) relationships. The edge-weighted graph visualisation (Figure 4-22) illustrates strong outgoing *provider-to-customer* relationships starting at Level 3 Communications Inc. Additional strong relationships seemed to be established between Bharti Airtel Ltd. and Level 3 Communications Inc. and between a set of Autonomous Systems of Bharti Airtel Ltd. These findings demonstrate that Bharti Airtel shows a bottleneck reliance on Bharti Airtel, their own *Internet Service Provider* over which they have BGP routing control, while also benefiting from additional upstream (possibly paid) *peer-to-peer* relationships.

Interestingly, Vodafone shows that most of the fused Autonomous System relationships are of *peer-to-peer* nature (2.35% of the total relationships), followed by *customer-to-provider* relationships (1.48%) and *provider-to-customer* (1.17%) ones. Moreover, the

edge-weighted graph visualisation illustrates that the strongest linkages were established between the China Education and Research Network and Vodafone India Ltd. (*customer-to-provider*). Vodafone also displays a number of undiscovered relationships between Vodafone India Ltd. and Vodafone's Cable and Wireless Worldwide Plc. subsidiary. Other fairly strong relationships were a *peer-to-peer* relationship between Cogent Communications and Cable and Wireless Worldwide Plc. and a *customer-to-provider* relationship between Wireless Worldwide Plc. and Telia Company AB. These findings indicate that Cable and Wireless Worldwide Plc. represents a bottleneck *Internet Service Provider* for the Vodafone network. However, Vodafone recently acquired their upstream Cable and Wireless Worldwide Plc. partner for approximately US\$ 1.9 bn, which makes them a subsidiary of the Vodafone corporation (Indian Express, 2012), potentially indicating *sibling* relationships (see Gao, 2001, p. 734). This means that Vodafone eliminated part of their reliance on this structural bottleneck, creating a favourable position compared to the upstream Internet markets of Aircel and Bharti Airtel. Nevertheless, Vodafone still seems structurally reliant Telia Company AB. The insights also indicate that the China Education and Research Network is reliant on Vodafone India Ltd. as a structural bottleneck for global connectivity purposes, given the *customer-to-provider* relationship.

As a result of these findings, we claim that Bharti Airtel and Vodafone have a competitive advantage over Aircel, given the 'elimination' of their structural upstream bottlenecks. However, these Autonomous System relationships might still create certain upstream internetworking dependencies for reaching the Internet periphery.

Moreover, being structurally reliant on some influential Autonomous Systems for upstream connectivity purposes along the global digital supply chain displays the presence of asymmetries in the centrality of our three analysed Tamil Nadu mobile broadband operator networks. By studying Internet Exchange Point (IXP) routing decisions, D'Ignazio and Giovannetti (2006) indicate the existence of such asymmetries. Our evidence shows that Bharti Airtel and Vodafone are dependent on their very own Autonomous Systems for routing purposes to the network cores, which is a beneficial situation. Notably, a large number of Autonomous System relationships remain uncovered from the CAIDA (2016b) AS-relationship dataset (see section 4.4 for details), displaying the benefit of using active Internet periphery measurements, as Faggiani et al. (2012) indicate. Moreover, we acknowledge that only small parts of the economic

relationships actually emerged. Most connectivity connections are still missing (see above). This prohibits us to provide a clear economic picture of all Autonomous System relationships in the three mobile broadband operator networks and leads us to state our findings cautiously. However, we can indicate that our exploratory approach to research using active Internet periphery measurements discovers some key upstream relationships amongst key Autonomous Systems with structural bottlenecks or network core features.

Due to our exploratory and elaborated findings, we do not reject Working Hypothesis WH1.2: ‘Studying the Tamil Nadu mobile broadband operators from an Internet-Periphery perspective indicates previously hidden upstream AS relationships’.

6.1.4 Summary Hierarchical Structuring

This dissertation found evidence that the three studied Tamil Nadu mobile broadband operators display features of a hierarchical upstream Internet market structure. These features seemed more apparent for those mobile broadband operators that employed stronger upstream *peer-to-peer* relationships for internetworking purposes than *customer-to-provider* or *provider-to-customer* relationships. Moreover, we indicated a new set of metrics for studying hierarchical Internet market structuring, based on the *Eigenvector Centrality* and the *Clustering Coefficient*. By using the *k-core decomposition* in Chapter 4 and the *Descriptive Statistics* in Chapter 5, we also identified a set of specific upstream Autonomous Systems that are key for the global connectivity of the three Tamil Nadu mobile broadband operators. Furthermore, by combining the *Complex Network Analysis* with the secondary CAIDA (2016b) AS-Relationship dataset in Chapter 4, we indicated previously hidden upstream Autonomous System relationships. However, our analysis will benefit from additional business relationships data.

6.2 Hierarchical Structuring and Affordability

Based on insights from the Literature Review (see section 2.5), it was hypothesised in Working Hypothesis WH2 that the ‘Tamil Nadu mobile broadband operators that show signs of a hierarchical upstream Internet market structure offer less affordable mobile broadband price plans to an end-user’. This hypothesis was explored through the *Statistical Network Analysis* developed in Chapter 5. This analysis discovered interesting relationships between the level of hierarchical structuring of our mobile broadband operator networks and affordability as measured in price per Megabyte for each given price plan. In detail, Model 3.1 showed that an **increase** in the level of hierarchical

structuring resulted in an **increase** in price per Megabyte. The price per Megabyte itself is calculated as a price in Indian Rupees (INR) divided by the promised data allowance (e.g. 1GB in a given price plan), obtained from GSMOutlook (2015a, 2015b, 2015c). Interestingly, Model 3.2 shows that an **increase** in the level of hierarchical structuring resulted in an **increase** in the price per Megabyte. Model 4, which was based on the *Weighted In-Degree* connectivity, showed that an **increase** in the level of hierarchical structuring, based on *Clustering Coefficient*, results in an **increase** in price per Megabyte. Notably, based on the analysis of the role of *Eigenvector Centrality*, Model 4 also showed that a **decrease** in the level of hierarchical structuring results in a **decrease** in price per Megabyte.

Hence, all findings point towards the fact that, a more hierarchically structured upstream Internet market of the three Tamil Nadu mobile broadband operator networks results in less ‘affordable’ end-user price plans (measured in price per Megabyte). However, such influences of the level of hierarchical network structuring on prices emerge from different perspectives, hence increasing the robustness of our results. In detail, we note that the level of hierarchical structuring, as captured by the *Clustering Coefficient*, reflects the influence of the bargaining power of highly connected Autonomous Systems on their neighbouring ones, which increases their costs and the final price for end-users. Moreover, the *Eigenvector Centrality* captures different aspects of hierarchical structuring, depending on whether it relates to incoming or outgoing connectivity. In detail, a steeper relationship between the *Eigenvector Centrality* and outgoing connectivity captures an increase in the level of hierarchical structuring. This is due to the strengthened bottleneck role played by central Autonomous Systems to reach many other Autonomous Systems in the upstream Internet. On the contrary, a steeper relationship between *Eigenvector Centrality* and incoming connectivity captures a lower level of hierarchical structuring. This is because it indicates that central, influential Autonomous Systems receive many direct incoming connections from neighbouring Autonomous System, providing them with central access to the rest of the Internet.

This differentiation concerning the effects on the hierarchical structuring of the two separate relationships between incoming or outgoing connectivity and *Eigenvector Centrality* represents a novel approach and provides new findings which, to the best of our knowledge, have not been explored in the research domain yet. Here, based on the internetworking hierarchical structuring, a mobile broadband operator’s upstream

interconnection costs clearly influence the price per Megabyte and hence the affordability of mobile broadband for end-users in the Internet periphery. Moreover, affordability of mobile broadband price plans by definition is one of the key determinants of mobile broadband adoption rates. Addressing hierarchical bottlenecks and their effects on affordability will help to address some of the key issues underlying the present levels of digital divide, both in our explored Tamil Nadu case studies and more generally in the Internet periphery.

Given this novel insight, we do not reject Working Hypothesis WH2: ‘Tamil Nadu mobile broadband operators that show signs of a hierarchical upstream Internet market structure offer less affordable mobile broadband price plans to an end-user’.

6.3 Hierarchical Structuring and Quality of Service

Alongside the identified effects of the hierarchical upstream Internet market structure of the three mobile broadband operators on affordability, we also infer through Working Hypothesis 3, WH3, that ‘Those Tamil Nadu mobile broadband operators that show signs of a hierarchical upstream Internet market structure provide a lower Quality of Service to an end-user’

This hypothesis was empirically explored in Chapter 5 through a standard correlation table where the results revealed a number of interesting Quality of Service implications. Both the download speed and the throughput show a highly negative correlation with the hierarchical structuring covariates. Interestingly, the drop rate of a data packet ($\leq 5\%$) positively correlates with all the covariates from Model 1 and Model 2, indicating the importance of certain Autonomous Systems to the operator networks. Our correlations reveal potential effects between a hierarchical structuring of the upstream Internet market (for each of our mobile broadband operator networks) and the Quality of Service delivered to end-users residing in the Internet periphery. A deeper analysis going beyond these correlations would require further research to add explanatory power to these initial indications, going beyond the scope of our dissertation.

Our results lead us to preliminarily reject Working Hypothesis 3 (WH3) that ‘Those Tamil Nadu mobile broadband operators that show signs of a hierarchical upstream Internet market structure provide a lower quality of service to an end-user’. Based on the elaborated indications, it seems that a higher hierarchical clustering of the mobile broadband operator networks neither lowers nor increases Quality-of-Service to an end-

user. Instead, it might have an impact on specific Quality of Service metrics demanded by the Telecom Regulatory Authority of India (TRAI). Further research should shed light on these indications.

6.4 Implications to Practice and Policy

6.4.1 Implications to Practice

Our evidence indicates the Tamil Nadu mobile broadband operators' reliance on Autonomous Systems belonging to large upstream *Internet Service Providers* (Tier-1 or Tier-2). These Autonomous Systems are identified as forming some structural upstream connectivity bottlenecks due to their traffic routing roles in the digital supply chain of the Internet. Depending on their bilateral routing relationships, such bottlenecks might become relevant whenever an operator's upstream partner uses the identified key upstream Autonomous Systems for global connectivity purposes. Given the strong presence of such business relationships, we assume that the established Autonomous System relationships represent key barriers to entering into the Tamil Nadu upstream Internet market. These barriers potentially resulting in downstream value-driven strategies informed by important connectivity relationships. However, due to the general organisation of a fixed set of mobile broadband providers per Indian Telecom Service Areas (GOV-IN, 2016b), the Tamil Nadu mobile broadband operators potentially do not face threats of new market entrants when considering only the set of existing relationships. Moreover, the value of beneficial upstream connectivity relationships might not be forwarded to end-users. On the contrary, end-users seem to pay for the mobile broadband operators' cost for international connectivity. Nevertheless, new market entrants into the analysed upstream Internet market do indeed face barriers of entry given the minimum-cost link rule that Kang and Gligor (2014, p.10) indicate. This seems to apply to the relevance of the set of interconnectivity business relationships for our three upstream operator networks. New entrants could, therefore, only differentiate themselves either through the price (such as Reliance Jio) or through added values such as added cybersecurity and / or analytics services, introducing quality differentiation into an otherwise commoditised market.

These indications are of particular relevance since these key Autonomous Systems seemed to be frequently traversed when reaching the destination ASes of our collected *traceroute* measurements. Moreover, the identified asymmetry of Autonomous System

centralities would make settlement-free upstream peering less likely. This is because those *Internet Service Providers* at the core of the operator networks naturally aim for paid downstream relationships while potentially arranging settlement-free *peer-to-peer* connections amongst themselves. Therefore, connectivity asymmetries might result in ‘unfair’ competition conditions and stronger bargaining powers of large *Internet Service Providers* towards downstream Autonomous Systems that are situated closer to the Internet periphery (Tier-2 and Tier-3). Moreover, one could consider that upstream paid transit relationships might result in a lower bargaining power for Tier-1 and Tier-2 *Internet Service Providers* due to their strategic competitor differentiation choices. However, this could lead to situations where the Tier-1 *Internet Service Providers* are the only ones that can offer end-to-end supply of digital services at the ‘best price’ because of their economies of scale, even though they seem to clearly profit from their powerful hierarchical positions.

On the other hand, choosing settlement-free upstream *peer-to-peer* relationships might result in more affordable price plans for the end-users. This is assuming that the connectivity quality is the same and the mobile broadband operator passes upstream connectivity savings onto them. However, if an upstream Autonomous System peers with a multitude of downstream operators, then the upstream AS of a given operator might become a connectivity bottleneck itself because the Autonomous System would have to increasingly exchange traffic with a multitude of downstream operators. Moreover, if all of these key relationships were of *peer-to-peer* nature, then the situation would force *Internet Service Providers* to search for alternative Business Models, including those profiting from the analytics and sales of end-user data.

Nevertheless, a mobile broadband operator might still aim to add new upstream routes. This would allow operators to either profit from a cheaper transit or settlement-free peering relationships, or to circumvent bottleneck Autonomous Systems or the bargaining powers of large *Internet Service Providers*. Adding upstream routes still might not be the most beneficial way to reach favourable conditions for all parties involved. This view is supported by the Braess paradox which describes a similar problem with car traffic-flows in a game-theoretic setting (Braess, 1969).

Additionally, the hierarchical structuring of our Tamil Nadu mobile broadband operator networks shows their structural dependency on directly connected (first and / or second hop) upstream partners, representing the structural bottlenecks. This dependency on

upstream Autonomous Systems also refers to situations such as: cybersecurity risks and attacks; infrastructure changes and maintenance; infrastructure fallouts triggered by natural incidents such as the recent Tamil Nadu Internet service fallouts (e.g. at Tata Communications) caused by Cyclone Vardah in 2016 (India Times, 2016); non-natural incidents; or policy- or business-driven anti Net-Neutrality efforts. Since February 2016, the Telecom Regulatory Authority of India prohibits *Internet Service Providers* to charge discriminatory prices for data services based on content (TRAI, 2016e, p.2).

On the other hand, an upstream reliance on structural bottleneck Autonomous Systems might also result in sustainable bilateral business partnerships (or acquisitions in the case of Vodafone indicates). This could allow mobile broadband operators to negotiate more beneficial connectivity contracts or pricing in the future (see for example the Netflix Net Neutrality disputes with telecommunications conglomerate Comcast (CNET, 2014; Wired, 2016b)). While such partnerships for discriminatory prices are also prohibited by TRAI (2016e), they are very hard to identify in the hidden upstream Internet market. Nevertheless, such a bottom-up *negotiating* power seems valuable for *Internet Service Providers* since upstream price per Megabyte on the digital supply chain should constantly decrease, creating necessary recurring contract and partnership re-negotiations. Well-structured upstream connectivity could result in beneficial traffic-shaping mechanisms. This would allow the Tamil Nadu mobile broadband operators to offer more efficient connections to end-users and provide a potential competitive advantage in the fight for Tamil Nadu market share. This seems to be the case for Vodafone who owns their first directly-linked upstream Autonomous Systems in the upstream Internet market through which Vodafone strongly connects to Tier-1 operators for global connectivity to the Internet periphery. This indicates that Vodafone is able to route local traffic almost completely through their own Autonomous Systems. Future studies have to shed light on this situation, while also being able to profit from fewer dependencies on other partners.

These implications also seem relevant considering the sheer amount of information generated by end-users' smartphone usage. New mobile services such as social networking or streaming of music or video content, which are especially sought after in Tamil Nadu (Fennell et. al., 2016), are expected to result in considerable growth of the Indian mobile data usage in upcoming years (Ericsson, 2016). This rise might worsen the position of structural bottleneck Autonomous Systems, potentially resulting in routing

and bandwidth issues to reach the Internet periphery, assuming that bandwidth resources and their routing are not sufficiently increased or managed. Given the identified price per Megabyte effects, we argue that it is crucial for (the Tamil Nadu) mobile broadband operators to understand their complex upstream internetworking dependencies. It allows operators to decide upon those strategic directions that add the most value to increasingly demanding shareholders and end-users in the Internet periphery. However, it is not only the Tamil Nadu mobile broadband operators and their upstream partnering Autonomous Systems who need to be prepared for this expected rise in mobile broadband usage, national government authorities such as the Telecom Regulatory Authority of India should also be prepared. The following section discusses the implications of our evidence on policy.

6.4.2 Implications to Policy

Despite the upstream Internet market structure usually being overlooked by competition regulatory authorities, we argue that setting the right political discussion and agenda on these issues is crucial. These can be set from three perspectives. Firstly, to establish a sustainable, efficient and fair connectivity environment for *Internet Service Providers* in the upstream Internet market from a mobile broadband provider perspective. Secondly, to bring an affordable, non-discriminatory, efficient and secure connectivity to the Internet periphery from an end-user perspective. Lastly, to set rules to ensure a fair competition within the upstream Internet market for existing players and new entrants.

For this purpose, it would be important for policymakers to consider our indicated competition regulation trade-off between end-user price plan affordability (as measured in price per Megabyte) and the hierarchical structuring of the upstream Internet market. This includes their bottlenecks and results in connectivity asymmetries, and market and bargaining powers of large agents and critical small providers connecting to the core of the operator networks. This is especially relevant given a recent consultation paper by the Telecom Regulatory Authority of India, asking for options to ensure ‘fair, reasonable and non-discriminatory terms and conditions of interconnection agreements’ between Internet Service Providers (TRAI, 2016f, p.27). Alongside our identified effects of mobile broadband connectivity on supply-side policies, these themes are also relevant to the current policy focus on demand-side policies to regulate open and safer use of the Internet as stated in the WDR (2016). Our work adds to two demand-side issues. First, the openness of the Internet relates to equality and fair competition of the upstream Internet

market, where supply-side connectivity plays a crucial role. Second, we feel cybersecurity is as high on the agenda as Autonomous Systems since connectivity bottleneck positions might be at a greater risk of being targeted for cyber-attacks, while being critical in connecting end-users to the Internet.

Most importantly, our findings are of particular importance to the policy because of Tamil Nadu's rural and gender-based mobile broadband access and affordability disparities, as indicated in the Literature Review (section 2.2). A recent study on Internet inclusiveness by the Economist Intelligence Unit (EIU, 2017) commissioned by Facebook's internet.org unit ranked India as the highest rated country which had appropriate policies in place for ensuring future connectivity. However, India ranked fairly low in the Affordability category (Rank 26). This category looks at the access cost being relative to income and competition in the Internet market. India ranks 22nd in terms of competition and at a low 51st place in terms of Internet inclusiveness due to prices. This is especially interesting considering our evidence from Chapter 5 (described in section 6.2. above). For those Indians that fall under the World Bank's poverty income measurement of US\$3.10/day, a 500MB mobile broadband price plan represents a very large portion of their average income (2.51% of GNI p.c., see ITU, 2015, p.136). Moreover, this reflects the issue of urban-rural per capita income disparities between Tamil Nadu districts (see section 2.2 and e.g. Selvabaskar et al., 2016). India also shows large gender disparities between men and women in terms of education, lack of income, and social attitudes towards technology (UN, 2014; WDR, 2016). Affordable and equal access to mobile broadband remains a key challenge for India (Broadband Commission, 2014). These income inequalities might result in a further drop in India's Affordability Drivers Index (ADI), where it ranked 31st in 2015 / 2016 (A4AI, 2016). A similar trend is also apparent in The Web Index of the World Wide Web Foundation (2014) where India ranked a low 58th place for Access and Affordability (ranked 48th for the cost of mobile broadband per capita income). Interestingly, India ranks 8th out of 140 countries in terms of affordability in the most recent Global Information Technology 2016 Report's Network Readiness Index (NRI) by the WEF (2017a, p.110; 2017b). However as discussed in Chapter 2, one reason for this surprisingly favourable and high ranking is the fact that the World Economic Forum does not include (prepaid) mobile broadband in the affordability metric (see WEF, 2017a, p.35). This shortcoming is clearly surprising given the importance of mobile broadband in both the developing and the developed nations. Improving affordability is considered to be helping to lift the Indian mobile broadband penetration

rate to 68% by the end of 2020, representing 330 million new subscribers (GSMA, 2016). Improving affordability would undoubtedly help India to increase their ranking in a number of the following development indexes while stopping the current negative trends in some of them.

First, more affordable price plans could help to increase India's ICT Development Index (IDI), especially after their recent drop to 138th ranking in 2016 (from 135th in 2015), while directly affecting the ICT Development Access and Use sub-indices, as described in the Literature Review (see section 2.2.4). These sub-indices metrics include:

- Mobile-cellular telephone subscriptions per 100 inhabitants
- Percentage of households with Internet access
- Percentage of individuals using the Internet
- Active mobile broadband subscriptions per 100 inhabitants.

Second, India could potentially increase their position in some metrics of the Network Readiness Index (WEF, 2017a):

- Accessibility of digital content (currently rank 93)
- Impact of ICT on access to basic services (currently rank 42)
- Internet users (currently rank 118)

Third, a more affordable access to mobile broadband Internet might help to achieve three of the 'Digital India' project visions of GOV-IN (2016):

- Digital infrastructure as a core utility to every citizen
- Governance and Services on demand
- Digital empowerment of citizens

Lastly, more affordable Tamil Nadu mobile broadband price plans would increase access to the mobile Internet by the less-favoured population, overcoming some factors of the digital divide, especially in the rural districts of Tamil Nadu. It would also help to drive the agenda for Sustainable Development Goals of the United Nations (United Nations, 2017). Goal 9, 'Industry, Innovation and Infrastructure', in particular is directly affected by the affordability of mobile broadband price plans. This goal includes the importance of Information and Communication Technology (ICT) infrastructure and the resulting service capabilities. The specific targets (measurable outcomes) of this goal include:

- ‘Develop quality, reliable, sustainable and resilient infrastructure, including regional and trans-border infrastructure, to support economic development and human well-being, with a focus on affordable and equitable access for all’
- ‘Significantly increase access to information and communications technology and strive to provide universal and affordable access to the Internet in the least developed countries by 2020’

Furthermore, significantly more affordable access to the information and communication infrastructure might also result in individuals having greater access to information and enhancing human capital. This is a strong driver for improving well-being, economic growth (GSMA, 2016), education and employment (Fennell et. al., 2016), inclusion (Broadband Commission, 2016), equality and social impact (WDR, 2016). The Broadband Commission for Sustainable Development states that affordable access to mobile broadband would lift millions of people out of poverty while contributing greatly to India’s Gross Domestic Product (Broadband Commission, 2016). Affordable access to mobile broadband Internet is necessary for the development of life-changing and innovative mobile solutions such as applications and services in the areas of Money and Banking, Governance, Agriculture, Education, and Health (Fennell et al., 2016).

Ultimately and despite all these promising outlooks, our insights indicate that reducing the affordability through competition regulations in the upstream Internet market might come at a cost. Our evidence shows that there is a competition regulation trade-off between affordable mobile broadband price plans (as measured in price per Megabyte) and a hierarchical structure of the upstream Internet market. This trade-off results in greater market and bargaining powers of Autonomous Systems belonging to large Tier-2 and Tier-1 *Internet Service Providers* that capture and secure barriers of entry.

Our case study evidence indicates that a more hierarchical upstream Internet market structure results in reduced affordability due to a higher price per Megabyte. Moreover, a more hierarchical upstream Internet market structure would allocate even more power to these Autonomous Systems, resulting in potentially ‘unfair’ connectivity asymmetries as identified by D’Ignazio and Giovannetti (2006). Additionally, a more hierarchical structuring could result in increasing Net Neutrality and price-bargaining issues as well as in stronger structural connectivity bottlenecks. This would pose a risk to the efficiency (given its anticipated growth in bandwidth needs) and affordability of the critical digital supply chain in India, which could ultimately end up widening the existing level of digital

divide. Such structural bottlenecks may also affect the demand side of the digital supply chain. They could endanger central agents who risk being targeted by cyber-attacks aiming to disrupt bandwidth and Quality of Service, as shown in a recent Distributed Denial of Service (DDoS) attack on Dyn, a company controlling Domain Name Systems (DNS) on the Internet. This attack resulted in services such as Twitter or Netflix being offline (The Guardian, 2016).

Our findings and results should be considered by both Indian and International policymakers as the presence and role played by larger *Internet Service Providers* with crucial bargaining and market power negatively correlates with the affordability of mobile broadband. Moreover, bypassing connectivity traffic of bottleneck *Internet Service Providers* whenever possible, especially when routing local traffic, would allow for more affordable mobile broadband price plans for end-users. As usual, nothing worth having comes easy.

7 CONTRIBUTIONS, LIMITATIONS, GENERALISABILITY AND RECOMMENDATIONS

7.1 Contributions to Knowledge

By using *Complex* and *Statistical Network Analysis*, based on active Internet periphery measurements, our exploratory case study concludes with a number of novel findings that offer distinctive evidence for the following contributions to knowledge:

- Our combination of *Descriptive*, *Complex Network*, *Graph Visualisation* and *Statistical Network Analysis* using active Internet periphery measurements and secondary data proved to be an extremely valuable methodological combination to explore non-trivial upstream Internet market structures and structural bottlenecks of mobile broadband operators. Given the lack of preliminary work in this multidisciplinary field of research, we add Giovannetti and Sigloch's (2015) pilot experiment by proposing a novel approach to study the mostly hidden nature of upstream Internet market structures from an Internet periphery perspective.
- Our Graph Visualisation Analysis showed that the Barabási-Albert Model B is a very suitable model to visually study network growth emergence of the upstream Internet access market for each of our case study's three mobile broadband operators. This adds to the work of Albert, Jeong and Barabási (1999) and Barabási and Albert (1999, 2002). Barabási and Albert (1999, p.7) show that preferential attachment mechanisms are useful for Business, Social and Transportation networks. The algorithm is also clearly useful to study network growth emergence of connectivity networks from an upstream Internet perspective. Moreover, our research showed that the applied k-core decomposition revealed the key operator network market agents at both levels of analysis (IP and Autonomous System Number granularity). Therefore, we add to the large-scale network studies of Alvarez-Hamelin et al. (2005a, 2005b, 2008). While Alvarez-Hamelin et al. (2005b) showed the general applicability to study natural hierarchical structures using Autonomous Systems, we can clearly indicate that this approach is useful from an Internet periphery perspective, using

crowdsourced end-user data.

- We provide the first case study to fuse primary upstream connectivity data from active Internet periphery measurements using Portolan (2015) with secondary CAIDA (2016a) AS-Rank and CAIDA (2016b) Autonomous System relationship data. This fusion revealed a number of previously undiscovered Autonomous System relationships in the upstream Internet markets of our three studied Tamil Nadu mobile broadband operators. These data fusions are very important and might help researchers at CAIDA to extend their Internet mapping efforts, including an Internet periphery perspective into existing data collection and analysis methods.
- Our specified *Statistical Network Analysis* provided robust confidence towards the existence of hierarchical upstream connectivity structures for all three of the studied Tamil Nadu mobile broadband operators. These contributions build on the findings of hierarchical upstream connectivity structures for B-Mobile in Bhutan by Giovannetti and Sigloch (2015). Hence, we indicate a possibility that other mobile broadband operator networks from varying countries are similarly structured.
- We show that connectivity scales not only on a negative exponent of the *Clustering Coefficient* as Vázquez, Pastor-Satorras and Vespignani (2002) indicate, but also in combination with a positive exponent of the *Eigenvector Centrality*, a metric that covers vertex influence in a network. Our research is the first to utilise this metric to study the presence of hierarchical upstream Internet market structures. Moreover, our research provides the first steps into an original exploration of the opposite roles (and interpretation) of the relations between *Eigenvector Centrality* and incoming and/or outgoing connectivity as indicators of the degree of hierarchical structuring and to identify the potential presence of bottlenecks in the analysed networks.
- By using these originally-devised metrics on originally collected data, we indicated that the three Tamil Nadu Mobile broadband operators are reliant on large (Tier-1 and Tier-2) *Internet Service Providers* for their global connectivity.

No other research of which we are aware indicated this reliance for the three Tamil Nadu mobile broadband operators, nor used similar data or our originally-developed methods elsewhere.

- Most importantly, we have exposed the relevant effect that a more hierarchical upstream Internet market structure is associated with higher prices (measured in price per Megabyte), leading to less affordable mobile broadband price plans for end-users in the Internet periphery. This critical new contribution should help interdisciplinary researchers as well as policymakers to measure and compare the global upstream Internet market structures and the end-user affordability as metrics for infrastructural access development.

7.2 Assessment of Working Hypotheses

Given the above stated contributions to knowledge, we conclude with a final assessment of our abducted *Working Hypotheses* while generating, through the best explanation of our evidence, one new Hypothesis (see Methodology section 3.3.2):

WH1: ‘The Tamil Nadu mobile broadband operators’ upstream Internet market structure displays features of a hierarchical ordering’.

Assessment: WH1 is not rejected.

WH1.1: ‘The Tamil Nadu mobile broadband operators rely on an identified set of specific Internet Service Providers for their upstream connectivity’.

Assessment: WH1.1 is not rejected.

WH1.2: ‘Studying the Tamil Nadu mobile broadband operators from an Internet-Periphery perspective indicates previously hidden upstream AS relationships’.

Assessment: WH1.2 is not rejected.

WH2: ‘Tamil Nadu mobile broadband operators that show signs of a hierarchical upstream Internet market structure offer less affordable mobile broadband price plans to an end-user’.

Assessment: WH2 is not rejected.

WH3: ‘Those Tamil Nadu mobile broadband operators that show signs of a hierarchical upstream Internet market structure provide a lower Quality of Service to an end-user’.

Assessment: WH3 is temporarily not rejected.

In conclusion, we note that further research should be performed to provide additional explanatory power for an understanding of the above-stated *Working Hypotheses*. Additionally, we are prompted to develop two new Hypotheses when considering the opposing effects of the relationships between *Eigenvector Centrality* and incoming and/or outgoing connectivity as indicators of the degree of hierarchical structuring and the potential presence of bottlenecks in the analysed networks studied in Chapter 5 and its detailed insights (see section 5.4.5). These Hypotheses represent a potential starting point for future research.

The first of such Hypotheses reflects on the role of incoming connectivity and when matched to high *Eigenvector Centrality* clearly indicates the presence of a Hub-like Autonomous Systems. This redirects incoming traffic originating from many different downstream Autonomous Systems and could be stated as:

H1: increased ‘Hub-like Autonomous System influence’ on incoming connectivity in the upstream Internet market results in more affordable price plans to end-users.’

The second new Hypothesis will instead focus on the role played by the relation between outgoing connectivity and *Eigenvector Centrality*. It reflects the bottleneck nature of certain Autonomous Systems that are critically central and relevant and are nearly an unavoidable bottleneck for the upstream routing of Internet traffic, originating from the Internet periphery. This would lead us to formulate the following Hypothesis for future work:

H2: increased ‘Bottleneck-like Autonomous System influence’ on outgoing connectivity in the upstream Internet market results in less affordable price plans to end-users.

7.3 Case Study Limitations

Our previous analysis provided interesting and novel results on the nature of the relationships between affordability and hierarchical structuring. Nevertheless, we acknowledge that our collected evidence faces a number of major limitations that impose caution about the impact and interpretation of our findings.

The first, and greatest, limitation of this case study's potential impact is the geographical and temporal nature and the resulting girth of our traceroute data, collected during our data collection campaign in Tamil Nadu, India. Thanks to UKIERI funding, we were able to organise a short 5-day data collection campaign in Tamil Nadu, India, while being hosted at the Indian Institute of Technology, Madras (IITM). Unfortunately, the available resources prevented us to conduct any long term, state-wide or cross-state studies. Such information would have provided us with a clearer overview of the state's present upstream Internet market structures. Instead, we reduced the impact of this limitation by focussing on a more specific case. In detail, we collected data for a culturally important (the holy pilgrimage route between Chennai and Kancheepuram) and very diverse (urban and more rural) part of Tamil Nadu. This was an important focus when looking at affordability differences in relation to the local upstream Internet access market structure.

Additionally, local issuing regulations prevented us from obtaining fully functional SIM cards for all four Tamil Nadu mobile broadband operators active at the time. This was because applicants were required to provide proof of residency and citizenship to the relevant authorities prior to being able to activate SIM cards. However, we were able to obtain three functioning local SIM cards, one for each of the three major mobile broadband operators at the time (Aircel, Bharti Airtel and Vodafone). These geographical, organisational, regulatory and temporal limitations clearly affected the overall generalisability of our research findings and contributions. Nevertheless, a number of valuable aspects of our research are transferable and will be able to sustain research efforts of other researchers in all related disciplines.

Another important limitation of our results stems from the transformation of the collected *Paris traceroute* data, obtained through the Portolan (2015) based active Internet Periphery measurements, to their associated Autonomous System Numbers using the secondary Maxmind (2015) GeoIP2 dataset. The fusion of these datasets have clear implications on our findings, which is discussed in depth in Chapters 4 and 5. To address this limitation, and to minimise the risk of ill data-transformations, the Maxmind (2015) GeoIP2 database was thoughtfully chosen as the main fusion source since it represents one of the most comprehensive collections of IP address ranges that any Autonomous System may incorporate. Moreover, these risks were further mitigated by drawing sample tests while comparing them with the fusion results of other credible sources, including Hurricane Electric (2016), UltraTools (2016) and Team Cymru (2016).

Another technical limitation of our collection campaign was the limited amount of *Paris traceroute* observations collected from the smartphone issued with a Bharti Airtel SIM card. Also, while being connected to the mobile broadband operator network, we experienced that the lower-end smartphones (Lava and Karbon) were unable to maintain a stable configuration of Portolan (2015). While this most likely reflects hardware issues, a more thorough preparation of the smartphones might have avoided such lack of data. We minimised the risk associated with a low number of observations by thoroughly preparing the smartphones, their issued SIM cards, as well as the Portolan (2015) Android application.

Lastly, a final limitation can be found in the automated, and hence not necessarily realistic, data collection of *traceroute* observation using the Portolan (2015) Android application. End-users most often access subjectively relevant, often local language content, rather than a randomly chosen set of target IP addresses, as set within the adopted version of the Portolan application. Additionally, large parts of traffic consumption on the Internet stems from troll activities, automated bots or Artificial Intelligences. All of these aspects are currently not accounted for in the Portolan (2015) Android application. We minimised this limitation by stating that our case study represents the actual usage of a tourist or commuter visiting the Chennai and Kancheepuram districts in Tamil Nadu (see section 3.3.4). In doing so, we mimicked the touristic or commuting behaviour while driving from the city of Chennai to the district of Kancheepuram to visit different historical temple sites.

7.4 Generalisability and Transferability

The limitations, especially concerning the geographical and temporal nature of our case study as described above, clearly show that this research is not entirely generalisable. Moreover, the population of our mobile broadband operator networks sample was rather small and did not cover all operators in the studied state of Tamil Nadu, India. Our strategic choice of this case study was to achieve a greater understanding of the upstream Internet market structures that are present in a critically important lower-middle income country such as India (see also Chapter 2). While our evidence clearly fulfilled this desire, we identify the main value is in our gained ability to test the Working Hypotheses abducted in section 2.5. Beyond that, our cross-sectional case study design, and especially the choice of mobile broadband operators, the short data collection duration as well as the restricted geographical coverage, distinctively limit the generalisability of our evidence.

Hence, our findings should be reflected upon with caution. Besides the limited generalisability, numerous findings and contributions that mainly stem from our applied methodology are transferable to research in related disciplines:

Our Graph Visualisation Analysis revealed that the Barabási-Albert Model B is the most suitable model to visually study network growth emergence of the upstream Internet access market. Moreover, it showed that the applied k-core decomposition indicated the key operator network market agents at both levels of analysis (IP and Autonomous System Number granularity). The k-core decomposition in particular reduces the need for intensive descriptive statistics in future research since it incorporates key structural aspects in the algorithm.

- By using statistical methods and graph plots, we showed that the three operator networks in this case study followed *Scale-Free Network* models.
- We revealed that connectivity not only scales on a negative exponent of the *Clustering Coefficient* (see Vázquez, Pastor-Satorras and Vespignani, 2002), but also in combination with a positive exponent of the *Eigenvector Centrality*, a metric that covers vertex influence in a network.
- By merging our collected upstream connectivity data with secondary CAIDA (2016a, 2016b) AS-Rank and AS-Relationship data, we provided an overview of the general econometric nature of the upstream Internet market structures for each of the three studied Tamil Nadu mobile broadband operator networks.
- Our two-stage econometric estimation process showed that a more hierarchical upstream Internet market structure leads to lower affordability (as measured in price per Megabyte). This finding provides a relevant antecedent for further testing in different geo-temporal contexts.

7.5 Recommendations for Future Research

Given the limitation of the potential impact of this dissertation, future research may incorporate the following measures to compensate for our identified shortcomings:

- The geographical and temporal limitations of our case study may be compensated by covering larger geographical areas (or even entire countries) during long-term case studies. Such efforts could be organised with wide research collaboration. Nevertheless, those efforts require sufficient organisation, resources and research funding. To overcome some of these costs, future research could further harness

the crowdsourcing approach by including a large number of actual end-users (with incentives) into the data collection process. Such an approach would also reduce the difficulties in obtaining and issuing local SIM cards. One could even partner with Internet Service Providers (e.g. mobile broadband operators), who could include Internet periphery measurements into their general connectivity provisioning.

- The high dependency on the secondary Maxmind (2015) Geo IP2 dataset could be reduced by a wider sharing and opening of data between data generators and collectors. An endeavour to collect which IP address ranges are allocated to which Autonomous Systems could be valuable. This would need collaboration between the five Regional Internet Registries (RIPE NCC for Europe, ARIN for America, APNIC for Asia-Pacific, LACNIC for Latin America and AfriNIC for Africa). Moreover, we encourage researchers to openly share obtained raw data of future studies. This Open Science approach would advocate comparability efforts while harnessing routes towards fruitful and sustainable Internet policies on a national and international level. Our collected traceroute dataset is openly available through the following link:

<https://www.doi.org/10.6084/m9.figshare.6839666.v4>

- Our case study experienced that lower-end smartphones faced difficulties to maintain a stable connection with Portolan (2015). Nevertheless, we still believe that future researchers should include lower-end smartphones into their measurements. These measurements should replicate actual end-user usage limitations, such as low signal antennas, advanced and reliable GPS facilities, and shorter battery life, wherever possible. The relevance of these lower-end smartphones is also obvious because the cost of a smartphone constitutes an additional element in defining affordability, which provides key access opportunities to the relevant socio-demographic end-user in any given country.
- To address the need for a clearer real-world picture of the upstream Internet market structure, Portolan (2015) or related applications could include connectivity measurements for local end-users' usage, bots and Artificial Intelligences into their Internet periphery measurement applications. One suggestion may include that end-users can browse the web while collecting

traceroute observations, although such an approach would most likely cause a number of ethical considerations given the tracking of actual usage data. Similarly, future measurement campaigns should be accompanied by concomitant end-user surveys. These are necessary to identify the average end-user behavioural patterns which would then be accurately replicated through the choice of final IP address destination by the Portolan application. This would provide a valuable approximation of the real-world upstream internetworking situation from an Internet periphery perspective, where both ‘bots’ and ‘humans’ create traffic and hence all the content-embodied value delivered through the digital supply chain. This is an increasingly sought-after demand of the end-user, given the increasing usage of contents requiring large-volume data, such as video or music streaming. Moreover, future researchers should cautiously prepare suitable smartphones for their data collection purposes. The data collection instructions in the Appendices may support this task.

Finally, we believe that measuring the structure of the (upstream) Internet (market) is a collective effort. Researchers in those domains should enhance collaborations to reveal the economic nature of the Internet. In particular, the sharing and fusion of collected datasets (e.g. from CAIDA, Telegeography and Internet periphery measurements) could provide a fruitful approach to increase the generation of knowledge in this discipline. Finally, we hope that this dissertation might contribute to initiating a policy debate on the requirements for equal access to the Internet.

7.6 Conclusions

The upstream Internet market structure is often a hidden component, particularly when observed from an Internet Periphery perspective. Yet it is still possible to explore some of its essential features that are necessary to gain a better understanding of the presence of elements that shape the market power potentially held by key *Internet Service Providers*, and its implications for the end-user affordability of mobile broadband price plans. Low affordability prevents the potentially desired diffusion of mobile broadband adoption and hence might hamper access to markets and information. These obstacles may limit the possibility of achieving substantial economic growth (GSMA, 2016), education and employment (Fennell et. al., 2016), inclusion (Broadband Commission, 2016), equality, and social impact (WDR, 2016) through wider access to mobile broadband Internet. Given the vast income disparities among the different Tamil Nadu

districts, this exploratory research covered a particularly sensitive topic for telecommunication authorities in emerging and developing countries. Based on a primary crowdsourced collection of *traceroute* connectivity data using active Internet Periphery measurement through the Portolan (2015), this case study synthesised an exploratory approach to research. The novel methodological integration of *Complex*, *Graph Visualisation* and *Statistical Network Analysis* provided a fruitful approach for holistically exploring the upstream Internet market structuring of three Tamil Nadu mobile broadband operator networks. Our results uncovered valuable evidence of hierarchical upstream Internet market structures for each of the three studied Tamil Nadu mobile broadband operators. Furthermore, we exposed the central internetworking importance of large Tier-1 *Internet Service Providers*, possibly having a substantial market and bargaining power in the global digital supply chain. Moreover, we discovered previously hidden Autonomous System relationships, indicating the value of collecting primary *traceroute* data using active Internet Periphery measurements. The most relevant findings of this dissertation were only achievable through the exploratory identification and novel application of often-neglected *Complex Network* metrics: the *Clustering Coefficient* and the *Eigenvector Centrality* and their relation to both incoming and outgoing connectivity. Lastly, our *Statistical Network Analysis* demonstrated that a more hierarchical upstream Internet market structure of mobile broadband operator networks decreases affordability (captured by the prices per Megabyte of the relevant mobile broadband operators' price plans). The implications of this evidence should lead policymakers to carefully consider the role played by larger *Internet Service Providers* with crucial bargaining and market powers. Their presence may negatively correlate with the affordability of mobile broadband if upstream network access conditions are only left to the unregulated game and are shaped by the relative strength and bargaining power among operators. Moreover, the increased possibility of bypassing *Internet Service Providers* that act as connectivity bottlenecks (e.g. through a successful establishment of more sustainable viable, local or periphery-based key infrastructure such as Internet Exchange Points or dedicated Research and Education Networks) appears to be a relevant policy tool in reducing the barriers of achieving Sustainable Development Goals in the Internet periphery.

8 REFERENCES

A

A4AI, 2016. A4AI 2015/2016 Affordability Report. *A4AI*, [online] Available at: <<http://a4ai.org/wp-content/uploads/2016/04/A4AI-2015-16-Affordability-Report.pdf>> [Accessed 25 Mar. 2017].

A4AI, 2017. A4AI 2017 Affordability Report. *A4AI*, [online] Available at: <<http://1e8q3q16vyc81g8l3h3md6q5f5e.wpengine.netdna-cdn.com/wp-content/uploads/2017/02/A4AI-2017-Affordability-Report.pdf>> [Accessed 12 Dec. 2017].

Agliardi, E. and Giovannetti, E., 1998. Morphogenesis of an institution on a lattice game. *Discrete Dynamics in Nature and Society*, 2(3), pp.209-13.

Ahmed, A., 2016. AIMS 2016 Workshop on Active Internet Measurements: Peering vs. Transit. *CAIDA*, [online] Available at: <https://www.caida.org/workshops/aims/1602/slides/aims1602_aahmed.pdf> [Accessed 6 Jan. 2017].

Alaettinoglu, C., 1996. Scalable router configuration for the Internet. *In Proc. IEEE IC3N*.

Albert, R., Jeong, H. and Barabási, A., 1999. *Internet: Diameter of the World-Wide Web. Growth dynamics of the world-wide web*. [online] University of Modena. Available at: <http://www.agentgroup.unimore.it/Zambonelli/didattica/cas/L9/Barabasi_Diameter.pdf> [Accessed 8 Oct. 2016].

Alvarez-Hamelin, J., Dall'Asta, L., Barrat, A. and Vespignani, A., 2005a. *Analysis and visualization of large scale networks using the k-core decomposition*. [online] University of Buenos Aires. Available at: <http://cnet.fi.uba.ar/ignacio.alvarez-hamelin/eccs/ECCS05_alvarez-hamelin.pdf> [Accessed 9 Mar. 2015].

Alvarez-Hamelin, J., Dall'Asta, L., Barrat, A. and Vespignani, A., 2005b. K-core decomposition: A tool for the visualization of large scale networks. *arXiv preprint cs/0504107*, [online] Available at: <<https://arxiv.org/pdf/cs/0504107.pdf>> [Accessed 23

Oct. 2016].

Alvarez-Hamelin, J., Dall'Asta, L., Barrat, A. and Vespignani, A., 2008. K-core decomposition of Internet graphs: Hierarchies, self-similarity and measurement biases. *Networks and Heterogeneous Media*, 3(2), pp.371-93.

APNIC, 2016 *The Internet routing registry*. [online] APNIC. Available at: <<https://www.apnic.net/manage-ip/apnic-services/routing-registry>> [Accessed 6 Nov. 2016].

Arc GIS, 2016. *Arc GIS Mapping Software*. [online] ArgGIS. Available at: <<http://www.arcgis.com/features/index.html>> [Accessed 8 Oct. 2016].

B

Bagchi, K., 2005. Factors contributing to global digital divide: Some empirical results. *Journal of Global Information Technology Management*, 8(3), pp.47-65.

Bak, P. and Chen, K., 1991. Self-Organized Criticality. *Scientific American*, 264(1), pp.46-53.

Ballester, C., Calvó-Armengol, A. and Zenou, Y., 2006. Who's who in networks. Wanted: The key player. *Econometrica*, 74(5), pp.1403-17.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O., 2013. The diffusion of microfinance. *Science*, 341(6144).

Barabási, A. and Albert, R., 1999. Emergence of scaling in random networks. *Science*, 286(5439), pp.509-12.

Barabási, A. and Albert, R., 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics: The American Physical Society*. [online] Available at: <http://www.barabasilab.com/pubs/CCNR-ALB_Publications/200201-30_RevModernPhys-StatisticalMech/200201-30_RevModernPhys-StatisticalMech.pdf> [Accessed 8 Oct. 2016].

Barabási Labs, 2013. *The Barabási-Albert Model*. [online] Barabási Labs. Available at: <http://barabasilab.neu.edu/networksciencebook/download/network_science_december_ch5_2013.pdf> [Accessed 30 Sep. 2016].

Barnett, G.A. and Park, H.W., 2012. Examining the international Internet using multiple measures: New methods for measuring the communication base of globalized cyberspace. *Quality and Quantity*, 48, pp.563-75.

Barrat, A., Barthélemy, M., Pastor-Satorras, R. and Vespignani, A., 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America* [online] Available at: <<http://arxiv.org/pdf/cond-mat/0311416.pdf>> [Accessed 26 Sep. 2016].

Bartosiak, 2012. *Complex generators plugin*. [online] Gephi marketplace. Available at: <https://marketplace.gephi.org/plugin/complex-generators/> [Accessed 15 Oct. 2015].

Bastian, M., Heymann, S. and Jacomy, M., 2009. *Gephi: An open source software for exploring and manipulating networks*. [online] AAI. Available at: <<http://www.aaai.org/ocs/index.php/ICWSM/09/paper/viewFile/154/1009/>> [Accessed 23 Oct. 2016].

Bavelas, A., 1950. Communication patterns in task oriented groups. *Journal of the Acoustical Society of America*, 22, pp. 271-82.

BCC News, 2012. *Vodafone agrees £1bn takeover of C&W Worldwide*. [online] BBC. Available at: <<http://www.bbc.com/news/business-17810568>> [Accessed 8 Jul. 2015].

Beauchamp, M.A., 1965. An improved index of centrality. *Behavioral Science*, 10, pp.161-3.

Bennett, L. and Norris, P., 2001. *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge, United Kingdom: Cambridge University Press.

Benzi, M., 2014. Centrality and Communicability measures in complex networks: *Mathematical and computational aspects, I*. [online] Available at: <<http://www.dsi.unive.it/issn/wp-content/uploads/2014/07/Benzi1.pdf>> [Accessed 19 Feb. 2017].

Berners-Lee, T., 2015. *Tim Berners Lee Quote*. [image online] izquotes. Available at: <<http://izquotes.com/quote/110066>> [Accessed 21 Oct. 2016].

Bliss, B. and West, D., 2007. *Introduction to graph theory*. 3rd ed. Harlow: Prentice Hall.

- Boccaletti, S., Latora, V., Chavez, M., Moreno, Y., and Hwang, D., 2006. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5), pp.175-308.
- Bollobás, B., 2001. *Random Graphs*. 2nd ed. Cambridge, United Kingdom: Cambridge University Press.
- Bonacich, P., 1983. Representations for homomorphisms. *Social Networks*, 5(2), pp. 173-92.
- Bonacich, P., 1987. Power and centrality: A family of measures. *American journal of sociology*, 92(5), pp. 1170-82.
- Bonacich, P., 2007. Some unique properties of eigenvector centrality. *Social Networks*, [online] Available at: <<https://pdfs.semanticscholar.org/a7b6/6df4aa6dfac4c513c9c32e8dd6cf1720accb.pdf>> [Accessed 5 Feb. 2017].
- Borgatti, S.P., 2005. Centrality and network flow. *Social Networks*, [online] Available at: <<http://www.analytictech.com/borgatti/papers/centflow.pdf>> [Accessed 5 Feb. 2017].
- Borgatti, S. P. and Li, X., 2009. On social network analysis in a supply chain context. *Journal of Supply Chain Management*, 45(2), pp.5-22.
- Boyles, D.R., 2006. Dewey's epistemology: An argument for warranted assertions, knowing, and meaningful classroom practice. *Educational Theory*, 56(1), pp.57-68.
- Braess, D., 1969. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung*, 12(1), pp.258-68.
- Brandes, U., 2001. A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology*, 25, pp.163-77.
- Bramoullé, Y., Kranton, R. and D'Amours, M., 2014. Strategic interaction and networks. *The American Economic Review*, 104(3), pp.898-930.
- Brath, R. and Jonker, D., 2015. Graph analysis and visualization: Discovering business opportunity in linked data. Hoboken, NJ, United States: John Wiley & Sons.
- Breusch, T. and Pagan, A., 1979. A simple test for Heteroscedasticity and random coefficient variation. *Econometrica*, 47(5), p.1287.

Broadband Commission, 2014. *Broadband and Sustainable Development*. [online] Broadband Commission. Available at: <<http://www.broadbandcommission.org/documents/BB-Davos-Report2014.pdf>> [Accessed 21 Mar. 2017].

Broadband Commission, 2016. *The state of Broadband 2016: Broadband Catalyzing sustainable development*. [online] Broadband Commission. Available at: <https://www.itu.int/dms_pub/itu-s/opb/pol/S-POL-BROADBAND.17-2016-PDF-E.pdf> [Accessed 7 Oct. 2016].

Brown, R. and Saunders, M., 2006. *Doing your Dissertation in business and management: The reality of researching and writing*. London: Sage Publications.

Bryman, A., 2012. *Social research methods*. 4th ed. New York: Oxford University Press.

Brynjolfsson, E., 2013. *The key to growth? Race with the machines*. [online] TED. Available at: <https://www.ted.com/talks/erik_brynjolfsson_the_key_to_growth_race_em_with_em_the_machines> [Accessed 21 Mar. 2017].

Burch, H. and Cheswick, B., 1999. Mapping the Internet. *Computer*, 32(4), pp.97-8.

C

Carmi, S., Havelin, S., Kirkpatrick, S., Shavitt, Y. and Shir, E., 2005. *Jellyfish and other interesting creatures of the Internet*. [online] Huji.ac.il Available at: <<http://www.cs.huji.ac.il/~kirk/JellyfishDimes.ppt>> [Accessed 23 Oct. 2016].

Cerf, V., 1990. *The Internet Activities Board – RFC1120*. [online] IETF. Available at: <<https://tools.ietf.org/html/rfc1160.html>> [Accessed 11 Mar. 2016].

Cerf, V. and Kahn, R., 1974. A protocol for packet network intercommunication. *ACM SIGCOMM Computer Communication Review*, 35(2), p.71.

Casas-Roma, J., 2015. *Plotting the coreness of a network with R and igraph*. [online] Available at: <<https://jcasasr.wordpress.com/2015/02/03/plotting-the-coreness-of-a-network-with-r-and-igraph/>> [Accessed 23 Oct. 2016].

Busch, J., Beiró, M. and Alvarez-Hamelin, I., 2011. Understanding edge Connectivity in the Internet through core decomposition. *Internet Mathematics*, [online] Available at: <<https://arxiv.org/pdf/0912.1424.pdf>> [Accessed 6 Mar. 2017].

CAIDA, 2015. *IPv4 and IPv6 AS core: Visualizing IPv4 and IPv6 Internet Topology at a Macroscopic scale*. [online] CAIDA. Available at: <http://www.caida.org/research/topology/as_core_network/#about> [Accessed 10 Jun. 2016].

CAIDA, 2016a. *AS rank: AS ranking*. [online] CAIDA. Available at: <<http://as-rank.caida.org>> [Accessed 27 Sep. 2016].

CAIDA, 2016b. *Inferred AS relationships Dataset*. [online] CAIDA. Available at: <http://www.caida.org/data/request_user_info_forms/as_relationships.xml> [Accessed 27 Sep. 2016].

CAIDA, 2016c. *Skitter AS links Dataset*. [online] CAIDA. Available at: <https://www.caida.org/data/active/skitter_aslinks_dataset.xml> [Accessed 9 Oct. 2016].

CAIDA-Ark, 2016. *Hosting an Ark monitor*. [online] CAIDA. Available at: <<http://www.caida.org/projects/ark/siteinfo.xml>> [Accessed 9 Oct. 2016].

Caldarelli, G., Marchetti, R. and Pietronero, L., 2000. The fractal properties of Internet. *Europhysics Letters (EPL)*, 52(4), pp.386-91.

Carpenter, B., 1996. *IETF RFC 1958 Architectural Principles of the Internet*. [online] IETF. Available at: <<https://www.ietf.org/rfc/rfc1958.txt>> [Accessed 10 Jun. 2016].

Castells, M., 2001. *The Internet galaxy: Reflections on the Internet, business, and society*. New York: Oxford University Press.

Cheswick, B., Burch, H. and Branigan, S., 2000. Mapping and visualizing the Internet. *In USENIX Annual Technical Conference*, pp. 1-12.

Choi, J. H., Barnett, G. A. and Chon, B. S., 2006. Comparing world city networks: a network analysis of Internet backbone and air transport intercity linkages. *Global Networks*, 6(1), pp. 81-99.

Choi, S., Galeotti, A. and Goyal, S., 2014. *Trading in networks: Theory and experiment*. Faculty of Economics, University of Cambridge, [online] Available at: <http://www-1v70.rz.uni-mannheim.de/material/intermediation22_full.pdf> [Accessed 10 Oct. 2016].

CNN Money, 2016. *900 million Indians can't get online. Here's why*. [online] CNN. Available at: <<http://money.cnn.com/2016/03/09/technology/india-internet-access/>>

[Accessed 24 Oct. 2016].

Conversiontools, 2012. Convert HTML table to comma-separated or tab-separated CSV file online - web and document conversion tools. [online] Available at: <<https://conversiontools.io>> [Accessed 20 Jan. 2016].

Cook, K. S., Emerson, R. M., Gillmore, M. R. and Yamagishi, T., 1983. The distribution of power in exchange networks: Theory and experimental results. *American journal of sociology*, 89(2), pp. 275-305.

Cook, R. and Weisberg, S., 1983. Diagnostics for Heteroscedasticity in regression. *Biometrika*, 70(1), p.1.

Cotten, S., Tashakkori, A. and Teddlie, C., 1999. Mixed methodology: Combining qualitative and quantitative approaches. *Contemporary Sociology*, 28(6), p.752.

Crovella, M. and Bestavros, A., 1996. Self-similarity in world wide web traffic. Proceedings of the 1996 ACM SIGMETRICS international conference on Measurement and modeling of computer systems - SIGMETRICS '96, [online] Available at: <<http://didattica.uniroma2.it/assets/uploads/corsi/144379/Crovella-Bestavros.pdf>> [Accessed 8 Oct. 2016].

D

Deichmann, J., Eshghi, A., Haughton, D., Masnghetti, M., Sayek, S. and Topi, H., 2006. Exploring break-points and interaction effects among predictors of the international digital divide. *Journal of Global Information Technology Management*, 9(4), pp.47-71.

Dall'Asta, L., Alvarez-Hamelin, I., Barrat, A., Vázquez, A. and Vespignani, A., 2005. Statistical theory of Internet exploration. *Physical Review E*, 71(3).

Dewey, J., 1941. Propositions, warranted Assertibility, and truth. *The Journal of Philosophy*, 38(7), p.169.

D'Ignazio, A. and Giovannetti, E., 2006. Antitrust analysis for the Internet upstream market: A BGP approach. *Faculty of Economics, University of Cambridge*, [online] Available at: <http://econrsa.org/papers/w_papers/wp37.pdf> [Accessed 10 Oct. 2016].

D'Ignazio, A. and Giovannetti, E., 2009. Asymmetry and discrimination in Internet

peering: evidence from the LINX. *International Journal of Industrial Organization*, 27(3), pp.441-48.

D'Ignazio, A. and Giovannetti, E., 2014. Continental differences in the clusters of integration: Empirical evidence from the digital commodities global supply chain networks. *International Journal of Production Economics*, 147, pp. 486-97.

DIMES, 2012. *The DIMES project* [online] Available at: <<http://www.netdimes.org/new/?q=node/65>> [Accessed 9 Oct. 2016].

Dimitropoulos, X., Krioukov, D., Fomenkov, M., Huffaker, B., Hyun, Y., Claffy, K. and Riley, G., 2007. AS relationships. *ACM SIGCOMM Computer Communication Review*, 37(1), p.29.

Dosi, G., 1982. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*, 11(3), pp.147-62.

E

Easley, D. and Kleinberg, J., 2010. *Networks, crowds, and markets reasoning about a highly connected world*. New York: Cambridge University Press.

Eckstein, H., 2000. *Case Study and Theory in Political Science*. London: SAGE Publications.

Economic Times, 2014. *Article: Investments in broadband need to be encouraged in India: Ericsson - the economic times*. [online] Economic Times. Available at: <http://articles.economictimes.indiatimes.com/2014-12-18/news/57196437_1_mobile-broadband-digital-india-broadband-infrastructure> [Accessed 12 Oct. 2016].

Economides, N., 1995. The economics of networks. *International journal of industrial organization*, 14(6), pp.673-99.

Eick, S., 1996. Aspects of network visualization. *IEEE computer graphics and applications*, 16(2), pp.69-72.

EIU, 2017. *EIU Inclusive Internet Index*. [online] EIU. Available at: <<https://theinclusiveinternet.eiu.com>> [Accessed 25 Mar. 2017].

Electronic Frontier Foundation, 2016. *Net neutrality*. [online] EFF. Available at: <<https://www.eff.org/de/issues/net-neutrality>> [Accessed 24 Oct. 2016].

Elliott, R., Moore, K. and Worren, N., 2002. When theories become tools: Toward a framework for pragmatic validity. *Human Relations*, 55(10), pp.1227-50.

Elliott, M., and Golub, B., 2013. A network approach to public goods. *In Proceedings of the fourteenth ACM conference on Electronic commerce*, pp. 377-8.

Erdős, P. and Rényi, A., 1959. On Random Graphs I. *Math. Debrecen*, 6, pp.290-7.

Ericsson, 2016. *Europe mobility report Appendix India*. [online] Ericsson. Available at: <<https://www.ericsson.com/res/docs/2016/mobility-report/emr-rina-june-2016.pdf>> [Accessed 7 Oct. 2016].

ERNET, 2016. *History of ERNET*. [online] ERNET. Available at: <<http://www.eis.ernet.in/about/abouternet.html>> [Accessed 7 Oct. 2016].

Estellés-Arolas, E. and González-Ladrón-De-Guevara, F., 2012. Towards an integrated crowdsourcing definition. *Journal of Information science*, 38(2), pp.189-200.

F

Faggiani, A., Gregori, E., Lenzini, L., Mainardi, S. and Vecchio, A., 2012. *On the feasibility of measuring the Internet through Smartphone-based Crowdsourcing*. [online] IIT CNR. Available at: <<http://portolanproject.iit.cnr.it/wp-content/uploads/2014/12/1569571021.pdf>> [Accessed 5 Feb. 2015].

Faggiani, A., Gregori, E., Lenzini, L., Luconi, V. and Vecchio, A., 2013. Lessons learned from the design, implementation, and management of a smartphone-based crowdsourcing system. *Proceedings of First International Workshop on Sensing and Big Data Mining*, [online] Available at: <<http://portolanproject.iit.cnr.it/wp-content/uploads/2014/12/lessons.pdf>> [Accessed 12 Dec. 2015].

Faggiani, A., Gregori, E., Lenzini, L., Luconi, V. and Vecchio, A., 2014a. Smartphone-based crowdsourcing for network monitoring: Opportunities, challenges, and a case study. *IEEE Commun. Mag.*, [online] Available at: <<http://portolanproject.iit.cnr.it/wp-content/uploads/2014/12/commag.pdf>> [Accessed 1 Feb. 2015].

- Faggiani, A., Gregori, E., Improta, A., Lenzini, L., Luconi, V. and Sani, L., 2014b. A study on traceroute potentiality in revealing the Internet AS-level topology. *2014 IFIP Networking Conference*, [online] Available at: <<http://portolanproject.iit.cnr.it/wp-content/uploads/2014/12/study.pdf>> [Accessed 25 Feb. 2017].
- Faloutsos, M., Faloutsos, P. and Faloutsos, C., 1999. On power-law relationships of the Internet topology. *ACM SIGCOMM Computer Communication Review*, 29(4), pp.251-62.
- Fay, D., Haddadi, H., Thomason, A., Moore, A., Mortier, R., Jamakovic, A., Uhlig, S. and Rio, M., 2010. Weighted spectral distribution for Internet Topology analysis: Theory and applications. *IEEE/ACM Transactions on Networking*, [online] Available at: <<http://mor1.github.io/publications/pdf/ton10-wsd.pdf>> [Accessed 6 Mar. 2017].
- Feldman, D. and Shavitt, Y., 2008. Automatic large-scale generation of internet pop level maps. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pp. 1-6.
- Fennell, S. and Arnot, M., 2007. Gender education and equality in a global context: Conceptual frameworks and policy perspectives. London, United Kingdom: Routledge.
- Fennell, S., Prashant, S., Charles, A., Ravindran, D., Kaur., P. and Jhunjhunwala, A., 2016. Internet Access for rural India. 27th European Regional Conference on the International Telecommunications Society (ITS), Cambridge, United Kingdom, 7th – 9th September 2016.
- Fershtman, C. and Gandai, N., 2011. Direct and indirect knowledge spillovers: the “social network” of open-source projects. *The RAND Journal of Economics*, 42(1), pp. 70-91.
- Flare Toolkit, 2010. *Data visualization for the web*. [online] Prefuse.org. Available at: <<http://flare.prefuse.org>> [Accessed 8 Oct. 2016].
- Floyd, S. and Paxson, V., 2001. Difficulties in simulating the Internet. *IEEE/ACM Transactions on Networking*, [online] Available at: <<http://icir.net/vern/papers/sim-difficulty.TON.2001.pdf>> [Accessed 8 Oct. 2016].
- Flyvbjerg, B., 2006. Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), pp.219-45.

Forbes, 2017. *Net Neutrality and The Unfortunate Politics Of Digital*. [online] *Forbes*. Available at: <<https://www.forbes.com/sites/steveandriole/2017/03/21/net-neutrality-the-unfortunate-politics-of-digital/#66f8edc61b4e>> [Accessed 17 Apr. 2017].

Foster, W., Goodman, S., McHenry, W., Press, L. and Wolcott, P., 2001. A Framework for Assessing the Global Diffusion of the Internet. *Journal of the AIS*, 2(6).

Freeman, L., 1977. A set of measures of centrality based on Betweenness. *Sociometry*, 40(1), p.35.

Freeman, L., 1978. Centrality in social networks conceptual clarification. *Social Networks*, 1(3), pp.215-39.

Freeman, L. C., Borgatti, S. P. and White, D. R., 1991. Centrality in valued graphs: A measure of betweenness based on network flow. *Social networks*, 13(2), pp.141-54.

Frieden, R., 2001. Does a Hierarchical Internet Necessitate Multilateral Intervention?. *North Carolina Journal of International Law*, 26(2).

Fruchterman, T. and Reingold, E., 1991. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11), pp.1129-64.

G

Gaertler, M. and Patrignani, M., 2004. Dynamic analysis of the autonomous system graph. *IPS 2004, International Workshop on Inter-domain Performance and Simulation*, [online] Available at: <<https://web.archive.org/web/20110722065122/http://www.dia.uniroma3.it/~patrigna/papers/files/ASGraphDynamicAnalysis.pdf>> [Accessed 6 Mar. 2017].

Galeotti, A. and Goyal, S., 2010. The Law of the Few. *American Economic Review*, 100(4), pp.1468-92.

Galvan, M. and Galvan, J., 2009. Writing literature reviews: A guide for students of the social and behavioral sciences. 4th edition. Glendale, CA: Pyrczak Pub.

Gao, L., 2001. On inferring autonomous system relationships in the Internet. *IEEE/ACM Transactions on Networking*, 9(6), pp.733-45.

Gao, L. and Rexford, J., 2000. Stable Internet routing without global coordination. *ACM*

SIGMETRICS Performance Evaluation Review, 28(1), pp.307-17.

Gao, L., Griffin, T. and Rexford, J., 2001. Inherently safe backup routing with BGP. *Proc. IEEE INFOCOM*, 1.

Garas, A., Argyrakis, P., Rozenblat, C., Tomassini, M. and Havlin, S., 2010. Worldwide spreading of economic crisis. *New Journal of Physics*, 12(11), p.113043.

Gehring, K. and Kishore G. K., 2008. Economic growth and income inequality in India. *Journal of Management*, 6(2), pp. 1-15.

Gephi, 2016. *The open graph Viz platform*. [online] Gephi.org. Available at: <<http://gephi.org>> [Accessed 2 Mar. 2014].

Gilbert, E., 1959. Random graphs. *Annals of Mathematical Statistics*, 30, pp.1141-4.

Giotsas, V. (2014). *Improving the accuracy of the Internet Cartography*. Ph.D. University College London.

Giovannetti, E., 2002. Interconnection, differentiation and bottlenecks in the Internet. *Information Economics and Policy*, 14(3), pp.385-404.

Giovannetti, E. and Sigloch, S., 2015. An Internet periphery study: Network centrality and clustering for mobile access in Bhutan. *Telecommunications Policy*, 39(7), pp.608-22.

Gladwell, M., 2015. *The tipping point*. 1st ed. London: Abacus.

Golub, B., & Jackson, M. O., 2010. Using selection bias to explain the observed structure of internet diffusions. *Proceedings of the National Academy of Sciences*, 107(24), pp.10833-6.

Gonzalez-Ladron-de-Guevara, F. and Estelles-Arolas, E., 2012. Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), pp.189-200.

Gorman, S. and Malecki, E., 2000. The networks of the Internet: An analysis of provider networks in the USA. *Telecommunications Policy*, [online] Available at: <<http://www.sciencedirect.com/science/article/pii/S0308596100000057>> [Accessed 19 Sep. 2016].

GOV-IN, 2016a. *Department of electronics & information technology*. [online] Government of India. Available at: <<http://digitalindia.gov.in>> [Accessed 12 Oct. 2016].

GOV-IN, 2016b. *Introduction to unified access services/cellular mobile services*. [online] Government of India. Available at: <<http://dot.gov.in/access-services/introduction-unified-access-servicescellular-mobile-services>> [Accessed 7 Oct. 2016].

Gov-IN, 2017. *Mobile Telephone Subscribers*. [online] Government of India. Available at: <<https://community.data.gov.in/mobile-telephone-subscribers-as-on-29-02-2016/>> [Accessed 5 Mar. 2017].

Goyal, S., 2009. *Connections: An Introduction to the Economics of Networks*. 2nd ed. Princeton, NJ: Princeton University Press.

Granovetter, M., 1985. Economic action and social structure: The problem of Embeddedness. *American Journal of Sociology*, 91(3), p.481.

Gregori, E., Lenzini, L., Luconi, V. and Vecchio, A., 2013. Sensing the Internet through crowdsourcing. *2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, [online] Available at: <<http://portolanproject.iit.cnr.it/wp-content/uploads/2014/12/p248-gregori.pdf>> [Accessed 9 Oct. 2016].

GSMA, 2016. *The Mobile Economy - India 2016*. [online] GSMA. Available at: <<http://www.gsma.com/mobileeconomy/india/>> [Accessed 21 Mar. 2017].

GSMOutlook, 2015a. Aircel online recharge, Aircel prepaid Tamil Nadu recharge plans ,Aircel prepaid Tamil Nadu Internet recharge, Aircel prepaid Tamil Nadu SMS packs, Aircel prepaid Tamil Nadu 2G packs, Aircel prepaid Tamil Nadu 3G packs, Aircel prepaid Tamil Nadu tariff pl. [online] GSMOutLook.com. Available at: <<http://www.gsmoutlook.com/offers-and-plans/aircel-prepaid-tamil-nadu-tariff-plans-internet-rechargesms-packs-3019.html>> [Accessed 16 Feb. 2017].

GSMOutlook, 2015b. Airtel prepaid Tamil Nadu recharge plans, Airtel prepaid Tamil Nadu Internet recharge, Airtel prepaid Tamil Nadu SMS packs, Airtel prepaid Tamil Nadu 3G recharge plans, Airtel prepaid Tamil Nadu tariff plans ,Internet recharge, SMS packs. [online] GSMOutLook.com. Available at: <<http://www.gsmoutlook.com/offers-and-plans/airtel-prepaid-tamil-nadu-tariff-plans-internet-rechargesms-packs-2509.html>>

[Accessed 16 Feb. 2016].

GSMOutlook, 2015c. BSNL Prepaid Tamil Nadu Recharge Plans ,BSNL Prepaid Tamil Nadu Internet Recharge,BSNL Prepaid Tamil Nadu SMS Packs, BSNL Prepaid Tamil Nadu 3G Recharge Plans,BSNL Prepaid Tamil Nadu Tariff Plans ,Internet Recharge,SMS Packs. [online] GSMOutLook.com. Available at: <<http://www.gsmoutlook.com/offers-and-plans/bsnl-prepaid-tamil-nadu-tariff-plans-internet-rechargesms-packs-2847.html>> [Accessed 10 May 2016].

GSMOutlook, 2015d. Vodafone prepaid Tamil Nadu recharge plans ,Vodafone prepaid Tamil Nadu Internet recharge, Vodafone prepaid Tamil Nadu SMS packs, Vodafone prepaid Tamil Nadu tariff plans ,Internet recharge, SMS packs. [online] GSMOutLook.com. Available at: <<http://www.gsmoutlook.com/offers-and-plans/vodafone-prepaid-tamil-nadu-tariff-plans-internet-rechargesms-packs-3337.html>> [Accessed 16 Feb. 2016].

Guillén, M. F., and Suárez, S.L., 2005. Explaining the global digital divide: Economic, political and sociological drivers of cross-national Internet use. *Social Forces*, 84(2), pp. 681-708.

H

Hall, J., 2013. Pragmatism, evidence, and mixed methods evaluation. *New Directions for Evaluation*, 2013(138), pp.15-26.

Harary, F., 1994.. *Graph theory*. 10th ed. Reading, MA: Addison-Wesley.

Hasani, F. and Mehdipour, E., 2015. Detecting infected Botnet machines by using the traffic behavior analysis. *Advances in Computer Science: an International Journal*, 4(5), pp.9-13.

Hawkinson, J. and Bates, T., 1996. Guidelines for creation, selection, and registration of an Autonomous System (AS) – RFC1930. *Internet Engineering Task Force*, [online] Available at: <<https://tools.ietf.org/html/rfc1930>> [Accessed 21 Oct. 2015].

Huberman, B. and Adamic, L., 1999. Growth dynamics of the world-wide web. *Nature*, 401(6749), p. 131.

Huffaker, B., Fomenkov, M. and Claffy, K., 2016. *CAIDA's Internet Topology data*

comparison. [online] CAIDA. Available at: <http://www.caida.org/research/topology/topo_comparison/> [Accessed 10 Dec. 2016].

Hurricane Electric, 2016. *Hurricane electric BGP Toolkit*. [online] BGP.he.net Available at: <<http://bgp.he.net>> [Accessed 01 Feb. 2016].

Huston, G., 1999. Interconnection, peering, and settlements. *proc. INET*. Vol. 9.

I

IANA, 2016. *Internet assigned numbers authority*. [online] Available at: <<https://www.iana.org>> [Accessed 5 Jul. 2016].

IBEF, 2016. *About Tamil Nadu: Tourism, industries, economy, agriculture, climate, geography*. [online] IBEF. Available at: <<http://www.ibef.org/states/tamil-nadu.aspx>> [Accessed 24 Oct. 2016].

Icompare, 2016a. *3G mobile Broadband providers in Chennai - compare data plans, tariffs and offers*. [online] [icompare.com](http://www.icompare.com). Available at: <<http://www.icompare.in/MobileBroadband/Providers?circle=Chennai>> [Accessed 7 Oct. 2016].

Icompare, 2016b. *3G mobile Broadband providers in Tamil Nadu (except Chennai) - compare data plans, tariffs and offers*. [online] [Icompare.com](http://www.icompare.com). Available at: <<http://www.icompare.in/MobileBroadband/Providers?circle=Tamil-Nadu>> [Accessed 7 Oct. 2016].

ICSI, 2016. *ICSI Netalyzer*. [online] University of Berkeley. Available at: <<http://netalyzer.icsi.berkeley.edu>> [Accessed 9 Oct. 2016].

Indian Express, 2012. *Vodafone to acquire CWW for \$1.9 bn*. [online] Indian Express. Available at: <<http://indianexpress.com/article/news-archive/web/vodafone-to-acquire-cww-for-1-9-bn/>> [Accessed 12 Nov. 2016].

India Times, 2016. *Vardah damages undersea cable, internet slows down*. [online] India Times. Available at: <<http://timesofindia.indiatimes.com/india/Vardah-damages-undersea-cable-internet-slows-down/articleshow/55967405.cms>> [Accessed 02 May. 2017].

Internet Live Stats, 2014. *India Internet users*. [online] internetlivestats.com. Available at: <<http://www.internetlivestats.com/internet-users/india/>> [Accessed 7 Oct. 2016].

Internet Live Stats, 2016. Internet users by country (2016). [online]] internetlivestats.com. Available at: <<http://www.internetlivestats.com/internet-users-by-country/>> [Accessed 7 Oct. 2016].

Internet Society, 2016. *A policy framework for enabling Internet access*. [online] Internet Society. Available at: <http://www.internetsociety.org/sites/default/files/Policy_Framework_for_an_Enabling_Environment_-_Stable.pdf> [Accessed 20 Oct. 2016].

Internet World Stats, 2016. *India Internet usage and telecommunications reports*. [online] Internet World Stats. Available at: <<http://www.internetworldstats.com/asia/in.htm>> [Accessed 7 Oct. 2016].

IRR, 2016. *IRR overview*. [online] Available at: <<http://www.irr.net/docs/overview.html>> [Accessed 6 Nov. 2016].

ITU, 2003. *ITU Internet reports 2003: Birth of Broadband*. [online] ITU. Available at: <<http://www.itu.int/osg/spu/publications/sales/birthofbroadband/index.html>> [Accessed 13 Oct. 2016].

ITU, 2007. *Discussion Paper International Internet interconnection*. [online] ITU. Available at: <https://www.itu.int/ITU-D/treg/Events/Seminars/GSR/GSR07/discussion_papers/Eric_lie_international_interconnection.pdf> [Accessed 16 Oct. 2016].

ITU, 2013a. *ICT Facts and Figures: The World in 2013*. [online] ITU. Available at: <<https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf>> [Accessed 18 Feb. 2015].

ITU (2013b). *Quality of Measurements of Broadband Internet*. [online] ITU. Available at: <http://www.itu.int/en/ITU-T/Workshops-and-Seminars/qos/201307/Documents/S5P2_Yvonne_Umutoni.pptx> [Accessed 10 Oct. 2016].

ITU, 2014. Economic and social impact of Broadband and development of digital agendas ITU/BDT regional economic and financial forum of telecommunications/ICTs for Latin America and the Caribbean. [online] ITU. Available at: <<https://www.itu.int/en/ITU->

D/Regulatory-Market/Documents/CostaRica/Presentations/Session3-1_Katz_Social%20Impact%20of%20BB.pdf> [Accessed 20 Oct. 2016].

ITU, 2015a. *ICT Facts and Figures 2015*. [online] ITU. Available at: <<https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf>> [Accessed 7 Oct. 2016].

ITU, 2015b. *2015 global ICT development index*. [online] ITU. Available at: <<http://www.itu.int/net4/ITU-D/idi/2015/>> [Accessed 11 Oct. 2016].

ITU, 2015c. *Measuring the information society report 2015*. [online] ITU. Available at: <<http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2015/MISR2015-w5.pdf>> [Accessed 11 Oct. 2016].

ITU, 2015d. *ITU ICT Eye India Profile 2015*. [online] ITU. Available at: <<https://www.itu.int/net4/itu-d/icteye/CountryProfileReport.aspx?countryID=113>> [Accessed 6 Nov. 2016].

ITU-D, 2016. *2016 Global ICT Development Index*. [online] ITU. Available at: <<https://www.itu.int/net4/ITU-D/idi/2016/#idi2016countrycard-tab&IND>> [Accessed 14 Mar. 2017].

ITU-D, 2017. *The ICT Development Index: conceptual framework and methodology*. [online] ITU. Available at: <<https://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2017/methodology.aspx>> [Accessed 6 Nov. 2017].

ITU-IDI, 2017. *The ICT Development Index (IDI): conceptual framework and methodology*. [online] ITU. Available at: <<https://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis/methodology.aspx>> [Accessed 13 May 2017].

J

Jackson, M., 2008. *Social and Economic Networks*. Oxford, United Kingdom: Princeton University Press.

Jacomy, M., Venturini, T., Heymann, S. and Bastian, M., 2014. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, 9(6).

Jhunjhunwala, A., Ramachandran, A. and Bandyopadhyay, A., 2004. n-Logue: The Story of a Rural Service Provider in India. *The Journal of Community Informatics*, [online] Available at: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.8848&rep=rep1&type=pdf#page=32>> [Accessed 16 May 2015].

Josephson, S. and Josephson, J., 1996. *Abductive inference: Computation, philosophy, technology*. Cambridge: Cambridge University Press.

K

Kagami, M., Tsuji, M. and Giovannetti, E., 2004. *Information Technology Policy and the Digital Divide*. Cheltenham, United Kingdom: Edward Elgar Publishing.

Kang, M. S. and Gligor, V. D., 2014. Routing bottlenecks in the internet: Causes, exploits, and countermeasures. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pp. 321-333.

Katz, M. and Shapiro, C., 1985. Network externalities, competition, and compatibility. *American Economic Review*, 75(3), pp.424-40.

Katz, R., Koutroumpis, P. and Callorda, F., 2013. The Latin American path towards digitization. *Info*, [online] Available at: <http://s3.amazonaws.com/academia.edu.documents/34032290/Digitization_index_v_final_Emerald.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1476993907&Signature=H31rLhkD1hMVVCJSaF0JVb6%2FOAY%3D&response-content-disposition=inline%3B%20filename%3DThe_Latin_American_path_towards_digitiza.pdf> [Accessed 20 Oct. 2016].

Kim, Y., Kelly, T. and Raja, S., 2010. *Building Broadband: Strategies and policies for the developing world*. [online] World Bank. Available at: <http://siteresources.worldbank.org/EXTINFORMATIONANDCOMMUNICATIONANDTECHNOLOGIES/Resources/282822-1208273252769/Building_broadband.pdf> [Accessed 26 Mar. 2016].

Kleinberg, J., 2000. The small-world phenomenon. *Proceedings of the thirty-second annual ACM symposium on Theory of computing - STOC '00*. [online] Available at: <<http://www.cs.cornell.edu/home/kleinber/swn.pdf>> [Accessed 26 Sep. 2016].

Knight, S., Nguyen, H. X., Falkner, N., Bowden, R. and Roughan, M., 2011. The internet

topology zoo. *IEEE Journal on Selected Areas in Communications*, 29(9), pp. 1765-75.

Knoke, D. and Yang, S., 2008. *Social network analysis*. Vol. 154. SAGE Publications: London.

Kogut, B., Shan, W., Walker, G., Eccles, R. and Nohria, N., 1992. The make-or-cooperate decision in the context of an industry network. *Networks and organizations*, pp.348-54.

Kolaczyk, E., 2009. *Statistical analysis of network data: Methods and models*. New York: Springer-Verlag New York.

Kuchar, 2012. *Layered layout - Gephi marketplace*. [online] Available at: <<https://marketplace.gephi.org/plugin/layered-layout/>> [Accessed 20 Feb. 2016].

Kumar, R., Raghavan, P., Rajalopagan, S. and Tomkins, A., 1999. Applications of linear algebra in information retrieval and hypertext analysis. *Proceedings of the 9th ACM Symposium on Principles of Database System*, p.1.

L

Lakhina, A., Byers, J., Crovella, M. and Matta, I., 2003. On the geographic location of internet resources. *IEEE Journal on Selected Areas in Communications*, 21(6), pp.934-48.

Liben-Nowell, D. and Kleinberg, J., 2008. Tracing information flow on a global scale using Internet chain-letter data. *Proceedings of the National Academy of Sciences*, 105(12), pp. 4633-8.

Luckie, M., Huffaker, B., Dhamdhere, A., Giotsas, V. and Claffy, K., 2013. AS relationships, customer cones, and validation. *Proceedings of the 2013 conference on Internet measurement conference*. [online] Available at: <<https://www.caida.org/~amogh/papers/asrank-IMC13.pdf>> [Accessed 18 Feb. 2015].

Luconi, V., Lenzini, L., Gregori, E. and Faggiani, A., 2015. Measuring the Internet Topology with Smartphones. *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdfunding of Big (Internet) Data*. [online] Available at: <<http://portolanproject.iit.cnr.it/wp-content/uploads/2014/11/measuring.pdf>> [Accessed 15 May 2016].

M

Mahadevan, P., Krioukov, D., Fomenkov, M., Dimitropoulos, X., and Vahdat, A., 2006. The Internet AS-level topology: three data sources and one definitive metric. *ACM SIGCOMM Computer Communication Review*, 36(1), pp.17-26.

Maxmind, 2015. *GeoIP2 Database*. [online] Available at: <https://dev.maxmind.com/geoip/geoip2/geoip2-csv-databases/#Locations_Files> [Accessed 20 Apr. 2015].

Medina, A., Matta, I. and Byers, J., 2000. On the origin of power-laws in Internet topologies. *ACM SIGCOMM Computer Communication Review*, 30(2), p.18.

Dorogovtsev, S., Goltsev, A. and Mendes, J., 2006. K -core organization of complex networks. *Physical Review Letters*, [online] Available at: <<https://arxiv.org/pdf/cond-mat/0509102.pdf>> [Accessed 23 Oct. 2016].

Minges, M., 2016. Digital dividends exploring the relationship between Broadband and economic growth exploring the relationship between Broadband and economic growth. [online] World Bank. Available at: <<http://pubdocs.worldbank.org/en/391452529895999/WDR16-BP-Exploring-the-Relationship-between-Broadband-and-Economic-Growth-Minges.pdf>> [Accessed 11 Oct. 2016].

Mintz, B. A. and Schwartz, M., 1985. *The power structure of American business*. Chicago : University of Chicago Press.

MIT, 2009. *Wide-Area Internet Routing*. [online] MIT Available at: <<http://web.mit.edu/6.033/www/papers/InterdomainRouting.pdf>> [Accessed 9 Jan. 2016].

Monge, P. R. and Contractor, N. S., 2003. *Theories of communication networks*. Oxford University Press, USA.

Morgan, D., 2014. Pragmatism as a paradigm for social research. *Qualitative Inquiry*, 20(8), pp.1045-53.

Mosaic Group, 1998. *Internet diffusion framework*. [online] Available at: <http://som.csudh.edu/fac/lpress/articles/acmfwk/acmfrwk.htm> [Accessed 9 Oct. 2016].

Murray, M. and Claffy, K., 2001. Measuring the Immeasurable: Global Internet Measurement Infrastructure. In *PAM—A workshop on Passive and Active Measurements*.

N

Nargundkar, R., 2003. *Marketing Research-Text & Cases*. 2nd ed. New Delhi: Tata McGraw-Hill Education.

Nefedov, N., 2013. *Complex Networks Analysis: Clustering Methods*. [online] Zurich: Federal Institute of Technology. Available at: <http://people.ee.ethz.ch/~nnefedov/can_lecture_01.pdf> [Accessed 11 Jun. 2015].

Netradar, 2016. *Netradar*. [online] Available at: <<http://Netradar.org>> [Accessed 24 Sep. 2016].

Newman, M., 2002. *Mathematics of networks*. Oxford : Oxford University Press.

Newman, M., 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23), pp.8577-82.

Newman, M., Barabási, A. and Watts, D., 2009. *The Structure and Dynamics of Networks*. Cambridge, United Kingdom: Cambridge University Press.

Newman, M. E. (2010). *Networks: An Introduction*. New York : Oxford University Press.

Noh, J. D., & Rieger, H., 2004. Random walks on complex networks. *Physical review letters*, 92(11), 118701.

Norton, W. B., 2011. *The Internet peering playbook: connecting to the core of the Internet*. DrPeering Press.

Nowotny, H., 2003. Democratising expertise and socially robust knowledge. *Science and Public Policy*, 30(3), pp.151-6.

O

OECD, 2008. *Broadband and the Economy*. [online] OECD. Available at: <<https://www.oecd.org/sti/ieconomy/40781696.pdf>> [Accessed 13 Oct. 2016].

Onnela, J. P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J. and Barabási, A. L., 2007. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18), pp.7332-6.

OpenSignal, 2016. *OpenSignal* [online] Available at: <<https://opensignal.com/about/>>

[Accessed 9 Oct. 2016].

OpenStreetMap, 2016. *OpenStreetMap*. [online] Available at: <<http://www.openstreetmap.org/#map=5/51.500/-0.100>> [Accessed 24 Sep. 2016].

Gregori, E., Lenzini, L. and Orsini, C., 2013. K-dense communities in the Internet AS-level topology graph. *Computer Networks*, [online] Available at: <<http://cat.inist.fr/?aModele=afficheN&cpsidt=27097039>> [Accessed 6 Mar. 2017].

P

Paris Traceroute, 2016. *Paris Traceroute* [online] Available at: <<https://paris-traceroute.net>> [Accessed 2 Jan. 2016].

Paul, J., 2002. Narrowing the digital divide: Initiatives undertaken by the association of south-east Asian nations (ASEAN). *Program: electronic library and information systems*, 36(1), pp.13-22.

Parvathamma, N., 2003. Digital divide in India: Need for correcting urban bias. *Information Technology and Libraries*, 22(1), p.35.

Pastor-Satorras, R., Vázquez, A. and Vespignani, A., 2001. Dynamical and correlation properties of the Internet. *Physical Review Letters*, 87(25).

Peirce, C.S., 1878. How to make our ideas clear. *Popular Science Monthly*, 12, pp.286-302.

Peirce, C. S., 1974. *Collected papers of Charles Sanders Peirce*. Vol. 2. Cambridge, United States: Harvard University Press.

Perer, A. and Shneiderman, B., 2006. Balancing systematic and flexible exploration of social networks. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), pp.693-700.

Pew Research, 2016. *Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies*. [online] Pew Research. Available at: <http://www.pewglobal.org/files/2016/02/pew_research_center_global_technology_report_final_february_22_2016.pdf> [Accessed 9 Jun. 2016].

Portolan, 2015. *PORTOLAN - mapping the Internet*. [online] IIT CNR. Available at:

<<http://portolanproject.iit.cnr.it>> [Accessed 9 Oct. 2016].

Powell, T. (2001). Competitive advantage: Logical and philosophical considerations. *Strategic Management Journal*, 22(9), pp.875-88.

Q

R

R, 2016. *R: The R Project for Statistical Computing*. [online] R-project.org. Available at: <<https://www.r-project.org>> [Accessed 5 Oct. 2016].

Reed, W., 2001. The Pareto, Zipf and other power-laws. *Economic Letters*, 74(1), pp.15-9.

Roberts, H., Larochelle, D., Faris, R. and Palfrey, J., 2011. Mapping local Internet control. *Computer Communications Workshop (Hyannis, CA, 2011)*, [online] Available at: <https://cyber.harvard.edu/netmaps/mlc_20110513.pdf> [Accessed 8 Oct. 2016].

Ruiz, J. and Barnett, G., 2015. Who owns the international Internet networks?. *The Journal of International Communication*, 21(1), pp.38-57.

Rubinov, M. and Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3), pp. 1059-69.

RIPE NCC, 2016a. *What is a remote route collector (RRC)?*. [online] RIPE NCC. Available at: <<https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/faq-ris/what-is-a-remote-route-collector-rrc>> [Accessed 6 Nov. 2016].

RIPE NCC, 2016b. *RIPE Atlas - RIPE network coordination centre*. [online] Available at: <<https://atlas.ripe.net>> [Accessed 9 Oct. 2016].

Rogers, E., 2001. The digital divide. Convergence: *The International Journal of Research into New Media Technologies*, 7(4), pp.96-111.

S

Sabidussi, G., 1966. The centrality index of a graph. *Psychometrika*, 31(4), pp.581-603.

Sandhusen, R., 2001. *Marketing*. 3rd ed. New York, United States: Barron's Educational Series Inc., U.S.

Saunders, M., Lewis, P., Thornhill, A. and Thorn, A., 2009. *Research methods for business students*. 5th ed. New York: Financial Times Prentice Hall.

Scott, J., Carrington, P. and Weihua, E., 2011. *The SAGE handbook of social network analysis*. London, United Kingdom: SAGE Publications.

Schneider, V. and Bauer, J., 2016. *Handbook on the economics of the Internet*. London, United Kingdom: Edward Elgar Publishing.

Schuster, S., Klamt, S., Weckwerth, W., Moldenhauer, F. and Pfeiffer, T., 2002. Use of network analysis of metabolic systems in bioengineering. *Bioprocess and Biosystems Engineering*, 24(6), pp.363-72.

Seidman, S., 1983. Network structure and minimum degree. *Social Networks*, 5(3), pp.269-87.

Selvabaskar, S., Alamelu, R., Amudha, R., Surulivel, S. T., and Motha, L. C. S., 2016. Understanding the 'Digital Divide'-An Investigation of Mobile and Mobile Service Usage among Rural and Urban Consumers in Delta Districts of Tamil Nadu. *Indian Journal of Science and Technology*, 9(27), pp.1-8.

Shavitt, Y. and Weinsberg, U., 2011. Quantifying the importance of vantage point distribution in Internet Topology mapping (extended version). *IEEE Journal on Selected Areas in Communications*, 29(9), pp.1837-47.

Shavitt, Y. and Zilberman, N., 2012. In: Geographical Internet PoP level maps. *International Workshop on Traffic Monitoring and Analysis 2012*.

Shea, T., Ariguzo, G. and White, D., 2006. Putting the World in the World Wide Web: The Globalization of the Internet. *International Journal of Business Information Systems*, 2(1), pp.75-98.

Shimbel, A., 1953. Structural parameters of communication networks. *The bulletin of mathematical biophysics*, 15(4), pp.501-7.

Siddhartha De, 2007. *Household Use of Information Technology, Australia, 2006-07*. [online] Australian Bureau of Statistics. Available at: <<http://www.abs.gov.au/Ausstats/abs@.nsf/0/acc2d18cc958bc7bca2568a9001393ae?OpenDocument>> [Accessed 13 Oct. 2016].

Siganos, G., Faloutsos, M., Faloutsos, P. and Faloutsos, C., 2003. Power-laws and the AS-level internet topology. *IEEE/ACM Transactions on Networking (TON)*, 11(4), pp.514-24.

Sigloch, S., Giovannetti, E., and Fennell, S., 2016. An exploratory network analysis of mobile broadband provider's infrastructure relationships in Tamil Nadu, India. *27th European Regional Conference of the International Telecommunications Society (ITS)*, Cambridge, United Kingdom, 7th – 9th September 2016, pp.1-22.

Sigloch, S., 2017. *WhatsApp Conversation with Eunbi Koh*. [chat] (Personal communication, 25 May 2017).

Simon, H., 1962. The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106(6), pp.467-82.

Singh, K., 2007. *Quantitative social research methods*. 2nd ed. Los Angeles: Sage Publications Pvt.

SLAC, 2001. *Stanford Linear Accelerator Center: Passive vs. Active Monitoring*. [online] Stanford University. Available at: <<https://www.slac.stanford.edu/comp/net/wan-mon/passive-vs-active.html>> [Accessed 17 Apr. 2017].

Sultana, 2016. *International Internet Connection (IIC) Cost: in aspect of Bangladesh*. [online] ITU. Available at: < <https://www.itu.int/en/ITU-T/Workshops-and-Seminars/bsg/201609/Documents/RSF%20for%20BSG%20India%20Bangladesh.pdf>> [Accessed 04 Jan. 2018].

Danesh, A., Trajkovic, L., Rubin, S. and Smith, M., 2001. Mapping the Internet. *Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, [online] Available at: <<http://www.cs.sfu.ca/~ljilja/papers/nafigs2001.pdf>> [Accessed 8 Oct. 2016].

Stake, R., 1995. *The art of case study research: Perspectives on practice*. 3rd ed. Thousand Oaks, CA: Sage Publications.

Stata, 2016. *Stata 13*. [online] Available at: <<https://www.stata.com/stata13/>> [Accessed 5 Jan. 2017].

Statista, 2016a. *Number of internet users in the Asia Pacific countries 2016* [online]

Available at: <<https://www.statista.com/statistics/265153/number-of-internet-users-in-the-asia-pacific-region/>> [Accessed 7 Oct. 2016].

Statista, 2016b. *Number of smartphone users in India 2013-2019*. [online] Available at: <<https://www.statista.com/statistics/467163/forecast-of-smartphone-users-in-india/>> [Accessed 7 Oct. 2016].

Stebbins, R., 2001. *Exploratory research in the social sciences*. Thousand Oaks, CA: Sage Publications.

Stephenson, K. and Zelen, M., 1989. Rethinking centrality: Methods and examples. *Social networks*, 11(1), pp. 1-37.

Streb, C., 2016. *SAGE Reference - exploratory case study*. [online] SAGE Publications. Available at: <<http://sk.sagepub.com/reference/casestudy/n139.xml>> [Accessed 1 Feb. 2016].

Strogatz, S., 2001. Exploring complex networks. *Nature*, 410(6825), pp.268-76.

Subramanian, L., Agarwal, S., Rexford, J. and Katz, R., 2002. Characterizing the Internet hierarchy from multiple vantage points. *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, 2, pp.1-11.

SWITCH, 2016. *Peering policy SWITCHlan*. [online] Available at: <<http://www.switch.ch/network/operations/peering-policy/>> [Accessed 16 Oct. 2016].

T

Tangmunarunkit, H., Govindan, R., Jamin, S., Shenker, S. and Willinger, W., 2002. Network topologies, power-laws, and hierarchy. *ACM SIGCOMM Computer Communication Review*, 32(1), p.76.

Team Cymru, 2016. *Team Cymru IP to ASN Lookup v1.0*. [online] Available at: <<http://asn.cymru.com>> [Accessed 12 Jan. 2016].

The Editorial Board, 2015. *Global threats to net neutrality*. [online] NY Times. Available at: <http://www.nytimes.com/2015/04/10/opinion/global-threats-to-net-neutrality.html?_r=0> [Accessed 24 Oct. 2016].

The Guardian, 2016. *DDoS attack that disrupted internet was largest of its kind in history*,

experts say. [online] The Guardian. Available at: <<https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet>> [Accessed on: 08 Nov. 2016].

The Ocp, 2016. *History of the Internet*. [online] Available at: <<http://www.thocp.net/reference/internet/internet1.htm>> [Accessed 15 Oct. 2016].

The Verge, 2017. *The FCC just killed net neutrality*. [online] Available at: <<https://www.theverge.com/2017/12/14/16776154/fcc-net-neutrality-vote-results-rules-repealed>> [Accessed 01 Feb. 2018].

The World Bank, 2016a. *World Bank Country Data: India*. [online] World Bank. Available at: <<http://data.worldbank.org/country/india>> [Accessed 7 Oct. 2016].

The World Bank, 2016b. *Data for lower middle income, India*. [online] World Bank. Available at: <<http://data.worldbank.org/?locations=XN-IN>> [Accessed 7 Oct. 2016].

The World Bank, 2016c. *Make Internet More Accessible, Affordable and Open to Accelerate Development, says New World Bank Report*. [online] World Bank. Available at: <<http://www.worldbank.org/en/news/press-release/2016/05/10/make-internet-more-accessible-affordable-open-accelerate-development-says-new-world-bank-report>> [Accessed 14 Jul. 2016].

Thomas, G., 2010. *How to do your case study: A guide for students and researchers*. London, United Kingdom: SAGE Publications.

Thompson, H. and Garbacz, C., 2011. Broadband Impacts on State GDP : Direct and Indirect Impacts. *17th Biennial Conference of the International Telecommunications Society*, pp.1-17.

TN-GOV-IN, 2014. *State of the Economy Chapter I*. [online] Government of Tamil Nadu. Available at: <<http://www.tn.gov.in/dear/State%20of%20the%20Economy.pdf>> [Accessed 24 Oct. 2016].

TN-GOV-IN, 2015. *State of the Economy Chapter II*. [online] Government of Tamil Nadu. Available at: <<http://www.tn.gov.in/dear/State%20Income.pdf>> [Accessed 25 Mar. 2017].

TRAI, 2010. *Recommendations on national Broadband plan*. [online] Telecom

Regulatory Authority of India. Available at: <<http://www.trai.gov.in/WriteReadData/Recommendation/Documents/Rcommendation81210.pdf>> [Accessed 7 Oct. 2016].

TRAI, 2014. *Press Release No. 83/2014*. [online] Telecom Regulatory Authority of India. Available at: <http://www.trai.gov.in/sites/default/files/PR%3D24122014_0.pdf> [Accessed 5 May 2016].

TRAI, 2015. *Annual Report 2014-2015*. [online] Telecom Regulatory Authority of India. Available at: <http://www.trai.gov.in/sites/default/files/TRAI_Annual_Report_English_16052016.pdf> [Accessed 5 May 2016].

TRAI, 2016a. *Letter to secretary DOT 24 may 2016*. [online] Telecom Regulatory Authority of India. Available at: <http://www.trai.gov.in/sites/default/files/Letter_to_Secretary_DOT_24_may_2016.pdf> [Accessed 13 Oct. 2016].

TRAI, 2016b. *Performance on Quality of Service of Wireless Data Services in June 2016*. [online] Telecom Regulatory Authority of India. Available at: <<http://www.trai.gov.in/sites/default/files/PMR%20Wireless%20ReportQE%20June%202016.pdf>> [Accessed 5 Oct. 2016].

TRAI, 2016c. *Performance Indicator Report July – September 2016*. [online] Telecom Regulatory Authority of India. Available at: <http://www.trai.gov.in/sites/default/files/Indicator_Reports_Ending_Sep_30122016.pdf> [Accessed 7 Dec. 2016].

TRAI, 2016d). *Telecom Regulatory Authority of India - Press Release 49/2016*. [online] Telecom Regulatory Authority of India. Available at: <http://www.trai.gov.in/sites/default/files/Press_Release_No.49_20_june_2016_Eng_0.pdf> [Accessed 7 Oct. 2016].

TRAI, 2016e. *Prohibition of Discriminatory Tariffs for Data Services Regulations*. [online] Telecom Regulatory Authority of India. Available at: <http://www.trai.gov.in/sites/default/files/Regulation_Data_Service.pdf> [Accessed 24 Oct. 2016].

TRAI, 2016f. *Consultation Paper on Review of the Regulatory Framework for Interconnection*. [online] Telecom Regulatory Authority of India. Available at: <http://www.trai.gov.in/sites/default/files/CP_on_Interconnection_21.11.2016.pdf> [Accessed 20 May 2017].

TRAI, 2017. *The Indian Telecom Services Performance Indicators Report October 2016*

– *December 2016*. [online] Telecom Regulatory Authority of India. Available at: <http://www.trai.gov.in/sites/default/files/Indicator_Reports_Dec_16_07042017.pdf> [Accessed 20 May 2017].

Tranos, E., 2013. *The geography of the internet: Cities, regions and internet infrastructure in Europe*. Cheltenham, United Kingdom: Edward Elgar Publishing.

Trochim, W., 2006. *Time in research*. [online] Available at: <<http://www.socialresearchmethods.net/kb/timedim.php>> [Accessed 10 May 2015].

U

UltraTools, 2016. *ASN lookup tool*. [online] Available at: <<https://www.ultratools.com/tools/asnInfo>> [Accessed 14 Oct. 2015].

United Nations, 2014. *United Nations e-Government Survey 2014*. [online] United Nations. Available at: <https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2014-Survey/E-Gov_Complete_Survey-2014.pdf> [Accessed 25 Mar. 2015].

United Nations, 2016. *United Nations statistics division- standard country and area codes classifications (M49)*. [online] United Nations. Available at: <<http://unstats.un.org/unsd/methods/m49/m49regin.htm#developed>> [Accessed 7 Oct. 2016].

United Nations, 2017. *United Nations Sustainable Development Goals*. [online] United Nations. Available at: <<http://www.un.org/sustainabledevelopment/sustainable-development-goals/>> [Accessed 8 Mar. 2017].

UNDP, 2016. *Sustainable Development Goals*. [online] UNDP. Available at: <<http://www.undp.org/content/undp/en/home/sustainable-development-goals.html>> [Accessed 11 Oct. 2016].

University of Adelaide, 2016. *The Internet Topology zoo*. [online] Available at: <<http://topology-zoo.org/index.html>> [Accessed 25 Sep. 2016].

Ushahidi, 2016. *Ushahidi*. [online] Available at: <<https://www.ushahidi.com>> [Accessed 24 Sep. 2016].

V

Van Dijk, J., 2006. Digital divide research, achievements and shortcomings. *Poetics*, [online] Available at: <<http://www.sciencedirect.com/science/article/pii/S0304422X06000167>> [Accessed 6 Nov. 2016].

Vázquez, A., Pastor-Satorras, R. and Vespignani, A., 2002. Large-scale topological and dynamical properties of the Internet. *Physical Review E*, 65(6).

Vega-Redondo, F., 2003. *Complex Social Networks*. Cambridge, United Kingdom: Cambridge University Press.

VSNL, 1996. *Internet Guide For VSNL's Gateway Internet Access Services (GIAS)*. [online] Available at: <<http://guide2net.net/bookweb/dnload/guide.pdf>> [Accessed 7 Oct. 2016].

W

Wang, X., Latapy, M. and Soria, M., 2012. Deciding on the Type of the Degree Distribution of a Graph from Traceroute-like Measurements. *International Journal of Computer Networks & Communications (IJCNC)*, 4(3).

Wasserman, S. and Faust, K., 1994. *Social network analysis: Methods and applications*. Cambridge, United Kingdom: Cambridge University Press.

Wasserman, S., and Galaskiewicz, J., 1994. *Advances in Social Network Analysis: Research from the Social and Behavioral Sciences*. Newbury Park, CA: SAGE Publications.

Waters, C., 2007. The nature and context of exploratory experimentation: An introduction to three case studies of exploratory research. *History and philosophy of the Life Sciences*, 29(3), pp.275-84.

Watts, D. and Strogatz, S., 1998. Collective dynamics of small world networks. *Nature*, 393(6684), pp.440-2.

Watts, D., 1999. Networks, dynamics, and the Small-World phenomenon. *American Journal of Sociology*, 105(2), pp.493-527.

Weinsberg, U. and Shavitt, Y., 2011. Quantifying the importance of vantage point

distribution in Internet Topology mapping. *IEEE Journal on Selected Areas in Communications*, 29(9), pp.1837-47.

Wiener, N., 2014. *Cybernetics or control and communication in the animal and the machine*. 1st ed. Cambridge, United States of America: M.I.T. Press.

Wikimedia Foundation, 2016. *History of Wikipedia*. [online] Available at: <<https://en.wikipedia.org/wiki/Wikipedia>> [Accessed on 8 Mar. 2016].

Willinger, W., Alderson, D. and Doyle, J., 2009. *Mathematics and the Internet: A source of enormous confusion and great potential*. [online] California Institute of Technology. Available at: <http://authors.library.caltech.edu/15631/1/Willinger2009p5466Notices_Amer._Math._Soc.pdf> [Accessed 28 Jun. 2016].

Willinger, W. and Roughan, M., 2013. *Internet Topology research Redux*. [online] University of Adelaide. Available at: <<http://www.maths.adelaide.edu.au/matthew.roughan/Papers/topology.pdf>> [Accessed 4 Jan. 2015].

Wired, 2006. *The Rise of Crowdsourcing*. [online] Wired. Available at: <<https://www.wired.com/2006/06/crowds/>> [Accessed 4 Apr. 2015].

Wolcott, P., Press, L., McHenry, W., Goodman, S. and Foster, W., 2001. A framework for assessing the global diffusion of the Internet. *Journal of the Association for Information Systems*, 2(1).

Wolcott, P., 2005. The Provision of Internet Services in India in Information Systems in Developing Countries. *Theory and Practice*, pp. 253-67.

Woodcock, B., 2003. *Internet Topology and Economics: How Supply and Demand Influence the Changing Shape of the Global Network*. [online] Packet Clearing House. Available at: <<http://www.pch.net/resources/papers/topology-and-economics/Topology-and-Economics.ppt>> [Accessed 16 Oct. 2016].

WEF, 2017a. *The Global Information Technology Report 2016*. Geneva, Switzerland: World Economic Forum.

WEF, 2017b. *WEF Network Readiness Index (NRI) in India*. [online] World Economic Forum Available at: <<http://reports.weforum.org/network-readiness-index/>> [Accessed 14 Mar. 2017].

WDR, 2016. *Digital Dividends*. [online] World Bank Available at: <<https://www.openknowledge.worldbank.org/handle/10986/23347>> [Accessed 25 Mar. 2017].

World Wide Web Foundation, 2014. *The Web Index*. [online] World Wide Web Foundation. Available at: <http://thewebindex.org/data/?indicator=ITU_O&country=IND> [Accessed 25 Mar. 2017].

World Wide Web Foundation, 2017. *The Web Index Net Neutrality*. [online] World Wide Web Foundation. Available at: <http://thewebindex.org/wp-content/themes/wixWordpress_Theme-develop/visualisations/neutrality-full.html> [Accessed 25 Mar. 2017].

X

XE, 2016. *INR - Indian rupee to USD United States Dollar converter*. [online] Available at: <<http://www.xe.com/currencyconverter/>> [Accessed 24 Oct. 2016].

Y

Yin, R.K., 2013. *Case Study Research: Design and Methods*. 5th ed. London, United Kingdom: SAGE Publications.

Yu, C., 2006. Abduction, deduction and induction. [online] Available at: <<http://www.creative-wisdom.com/teaching/WBI/abduction5.pdf>> [Accessed 22 Sep. 2016].

Z

9 APPENDICES

PUBLICATIONS, CONFERENCE PRESENTATIONS AND ABSTRACTS	333
TAMIL NADU MOBILE BROADBAND OPERATOR PRICE PLANS.....	339
KEY CHARACTERISTICS OF CASE STUDIES	343
DATA COLLECTION INSTRUCTIONS	344
ETHICAL CONSIDERATIONS, RELIABILITY, VALIDITY AND GENERALISABILITY.....	346
NUMBER OF OBSERVATIONS BY ORIGINATING AUTONOMOUS SYSTEM	348
POWER-LAW DEGREE DISTRIBUTIONS USING R	350
K-CORE DECOMPOSITION USING R	351
AS NUMBERS AND ORGANISATIONAL NAMES	353
AUTONOMOUS SYSTEM ROUTE PROPAGATIONS	354
STATA DO-FILES.....	358
ADDITIONAL GRAPH VISUALISATIONS	376

Publications, Conference Presentations and Abstracts

Giovannetti, E. and Sigloch, S., 2015. An Internet periphery study: Network centrality and clustering for mobile access in Bhutan. *Telecommunications Policy*, 39(7), pp.608-22.

Available at: <http://dx.doi.org/10.1016/j.telpol.2014.11.006>

[full article redacted due to copyright]

Sigloch, S., Giovannetti, E., and Fennell, S., 2016. An exploratory network analysis of mobile broadband provider's infrastructure relationships in Tamil Nadu, India. *27th European Regional Conference of the International Telecommunications Society (ITS)*, Cambridge, United Kingdom, 7th – 9th September 2016, pp.1-22.

[full article redacted due to copyright]

Available at: <http://hdl.handle.net/10419/148705>

UGC-UKIERI Workshop on Infrastructural Networks: Implications for Innovation in Rural Development. Indian Institute of Technology Madras, Chennai, India. 12th January 2016

**Mobile Internet Access and emerging bottlenecks:
Assessing the potential for shared electronic
infrastructure.**

UGC-UKIERI Workshop on Infrastructural Networks: Implications
for Innovation in Rural Development.
Indian Institute of Technology Madras, Chennai, India.
12th January 2016

27th European Regional Conference of the International Telecommunications Society.
The evolution of the north-south telecommunications divide: The role for Europe.
Cambridge, United Kingdom. 7th - 9th September 2016

**An exploratory network analysis of mobile broadband
provider's infrastructure relationships in Tamil Nadu, India.**

27th Regional Conference of the International Telecommunications Society
The evolution of the north-south telecommunications divide: The role for Europe.
Cambridge, United Kingdom.
7th - 9th September 2016

ABSTRACT: Sigloch, S., Giovannetti, E. and Fennell, S., 2017. 28th European Regional Conference of the International Telecommunications Society. Competition and Regulation in the Information Age. Passau, Germany. 30th July – 2nd August 2016.

28th European Regional Conference of the International Telecommunications Society (ITS):
Competition and Regulation in the Information Age
Passau, Germany, 30th July – 2nd August 2017.

**Upstream Internet Market Structures and Mobile Broadband Affordability,
A competition regulation trade-off in Tamil Nadu**

Sebastian Sigloch¹, Prof. Emanuele Giovannetti², Dr. Shailaja Fennell³

¹PhD Student, Anglia Ruskin University, Cambridge, United Kingdom

²Professor of Economics, Anglia Ruskin University, Cambridge UK

³Lecturer, Centre for Development Studies, University of Cambridge, Cambridge UK

Tamil Nadu faces a substantial digital divide. A recent Economist Intelligence Unit study on Internet inclusiveness ranked India as number one country regarding having appropriate policies for ensuring connectivity in place. This rank is achieved mostly due to the establishment of a digital identification, as well as a recent \$1.5 billion investment to bring some 250,000 villages to the Internet by 2018. Overall, however, India is only ranked 36th out of 75 due to the lower ranking in other category scores. Here, India ranked fairly low in the affordability category, which looks at the access cost about income and competition in the Internet market. While India ranked place 22 / 75 regarding competition, their price rank achieved a very low 51st rank amongst the studied countries. Moreover, India ranks eight / 139 for affordability in the most recent Global Information Technology Report Network Readiness Index (NRI) of the World Economic Forum (2016), a report that however does not include the importance of mobile broadband affordability.

For those Indians that fall below the World Bank's poverty line, pegged at income US\$3.10/day, a 500MB mobile broadband price plan represents a staggering 12% of their average income (ITU, 2015). This reflects the issue of urban – rural disparities. The South-East Indian state of Tamil Nadu is a special showcase at hand. While the state shows an above average per capita income, it displays a great digital divide between urban and rural districts. The rural district of Ariyalur e.g. fell, with a low US\$258 annual per capita income, far below the poverty line in 2013, see TN-GOV-IN (2015). Moreover, India faces great gender disparities between men and women about education, lack of income generation opportunities, and social attitudes towards technology (UN, 2014). Affordable and equal access to mobile broadband remains, therefore, a key challenge (Broadband Commission, 2014). A more affordable access to mobile broadband Internet might help to achieve the following three goals of the 'Digital India' visions as stated in GOV-IN (2016): i) Digital infrastructure as a core utility to every citizen, ii) Governance and Services on demand and iii) Digital empowerment of citizens. However, mobile broadband operators are reliant on their upstream partnering Internet Service Providers (ISPs) for reaching connectivity destinations and affordability targets. These upstream partners are linked either through settlement-fee based transit, or free peering, relationships. Hence, we infer the following research question: 'Does a more hierarchical upstream Internet structure, whose agents are

10th Anglia Ruskin University Annual Student Conference. June 2016. Cambridge, United Kingdom.

ANGLIA RUSKIN UNIVERSITY

ABSTRACT

LORD ASHCROFT INTERNATIONAL BUSINESS SCHOOL

DOCTOR OF PHILOSOPHY

MOBILE INTERNET ACCESS AND EMERGING BOTTLENECKS: ASSESSING
THE POTENTIAL FOR SHARED ICT INFRASTRUCTURE. EMPIRICAL
EVIDENCE FROM TAMIL NADU, INDIA.

SEBASTIAN SIGLOCH
(Post Confirmation of Candidature)

June 2016

This study analyses the emergent network features of the upstream connectivity structure, as being the data connection structure from an end-user's device to a predefined destination, for three mobile Internet Service Providers (ISPs) in the area included between the city of Kancheepuram and Chennai in Tamil Nadu, India, from an original end-users perspective. This perspective is based on capturing a large primary Paris Traceroute dataset, as being the most precise tool for following data packets along their route from source to destination. The data has been collected in February 2015, using the crowdsourcing-based Portolan Project Network Sensing Architecture, a mobile application being installed on three Android smart phones.

Internet Service Providers traditionally rely on Border Gateway Protocols (BGPs) for (paid) interconnection purposes between themselves and other ISPs, whereas BGPs represent the available connections, network policies and a set of rules for managing the exchange of information on the Internet. Those protocols are organised in so-called BGP routing tables that manage default interconnections. Adding this study's end-user perspective to the existing BGP routing tables reveals a more complete picture of the underlying network topology / structure and therefore co-operations among ISPs. The data were then used to conduct an Internet Periphery Analysis, pointing towards the roles of existing Internet traffic peering (exchange) agreements among the studied ISPs, and focusing on the role played by International Exchange Points (IXPs), a key electronic infrastructure for achieving efficient worldwide connections.

The analysis identifies the emergence of structural bottlenecks, the effectiveness of upstream ISP competition and the role of IXPs in providing a more widely distributed network access structure and proposes to further study the usage of settlement free peering in fair competition, the preconditions for an increase in Quality of Service to the respective end-users and fair service pricing in the Regional Mobile Broadband Market.

Key words: Mobile Broadband, Internet Topology, Network Analysis, ICT4D.

Tamil Nadu Mobile Broadband Operator Price Plans

Aircel price plans 24th January 2015			
#	Price in INR	Validity in days	Description as per GSMOutlook (2015a)
1	25	5	200MB 2G data -5Days
2	29	3	100 MB 3G & 2G Data
3	43	7	150 MB 3G & 2G Data
4	47	10	300MB 2G data-10Days
5	67	28	400 MB 3G & 2G Data + EXTRA USAGE Charges 25p/MB
6	98	14	500MB 2G Data
7	101	5	1 GB 3G & 2G Data
8	128	28	750 MB 3G & 2G Data + EXTRA USAGE Charges 25p/MB
9	148	28	1GB Data
10	193	28	1.8 GB Data
11	195	28	Unlimited 2G Data (2GB Highspeed). Thereafter Speed Throttled to 64 kbps
12	198	28	1.5 GB 3G & 2G Data
13	209	1	Unlimited (3 GB FUP) 3G&2G Data
14	255	28	Unlimited (1.5 GB FUP) 3G&2G Data
15	298	28	Unlimited (2 GB FUP) 3G&2G Data
16	309	1	Unlimited (5 GB FUP) 3G&2G Data
17	393	28	Unlimited (3.6 GB FUP)
18	399	28	Unlimited (3 GB FUP) 3G&2G Data
19	509	1	Unlimited (10 GB FUP) 3G&2G Data
20	693	28	Unlimited (7.2 GB FUP)
21	697	28	Unlimited (6 GB FUP) 3G&2G Data

22	993	28	Unlimited 12 GB FUP
23	997	28	Unlimited Data (10GB Highspeed). Thereafter Speed Throttled to 64 kbps
24	1397	28	Unlimited Data (15GB Highspeed). Thereafter Speed Throttled to 64 kbps
Key GB: Gigabyte INR: Indian Rupees MB: Megabyte			

Table 9-1: Aircel price plans 24th January 2015, Source: GSMOutlook (2015a).

Bharti Airtel price plans 24 th January 2015			
#	Price in INR	Validity in days	Description as per GSMOutlook (2015b)
1	8	1	25 MB 3G
2	28	3	100 MB 3G
3	45	5	150 MB 3G
4	101	14	400 MB 3G
5	127	14	650 MB 3G
6	197	28	1GB MB 3G
7	225	28	1 GB 3G + 150 MB Facebook
8	254	28	1 GB 3G + 150 MB Facebook + 200 MB WhatsApp
9	255	28	1.25 GB 3G
10	451	28	2.5 GB 3G
11	751	28	3G Unlimited, FUP 5 GB
12	955	28	3G Unlimited, FUP 7 GB
13	1298	60	6 GB 3G
14	1555	28	3G Unlimited, FUP 12 GB
15	2251	90	12 GB 3G
Key			

GB: Gigabyte
 INR: Indian Rupees
 MB: Megabyte

Table 9-2: Bharti Airtel price plans 24th January 2015, Source: GSMOutlook (2015b).

BSNL price plans 19 th January 2015			
#	Price in INR	Validity in days	Description as per GSMOutlook (2015c)
1	14	2	90 MB 2G / 3G Data
2	35	7	200 MB 2G / 3G Data
3	78	14	450 MB 2G / 3G Data
4	96	19	650 MB 2G / 3G Data
5	139	20	1 GB 2G / 3G Data
6	155	28	1 GB 2G / 3G Data usages Free with no speed restriction for Data, After free usages data will be charged at 2p / 10kb.
7	176	30	1 GB 2G / 3G Data
8	231	30	1.2 GB 2G / 3G Data (Offer valid up to 27 Dec 2014)
9	253	30	2GB 2G / 3G Data
10	561	30	5GB 2G / 3G Data
11	821	60	7GB 2G / 3G Data
12	1011	30	10GB 2G / 3G Data
13	1949	69	20GB 2G / 3G Data
Key GB: Gigabyte INR: Indian Rupees MB: Megabyte			

Table 9-3: BSNL price plans 19th January 2015, Source: GSMOutlook (2015c).

Vodafone price plans 25th January 2015.			
#	Price in INR	Validity in days	Description as per GSMOutlook (2015d)
1	8	1	Enjoy 25 MB Free 3G Internet usage on your Mobile or PC. Post Free usage 4P/10KB on Vodafone Live & 4P/10KB on Vodafone Mobile Connect
2	17	3	100 MB 2G data browsing
3	25	5	Enjoy 125 MB free 2G mobile internet
4	45	7	Enjoy 150 MB Free 3G Internet usage on your Mobile or PC. Post Free usage 4P/10KB on Vodafone Live & 4P/10KB on Vodafone Mobile Connect
5	49	7	Enjoy 250 MB free 2G mobile internet
6	98	14	Enjoy 500 MB free 2G mobile internet
7	102	14	400 MB 3G data browsing – Mobile Internet
8	124	21	Enjoy 650 MB free 2G mobile internet
9	148	28	Enjoy 1GB 2G free mobile internet
10	195	28	Enjoy 1GB 3G free mobile internet
11	199	28	Enjoy unlimited mobile internet (2GB Fair usage policy)
<p>Key</p> <p>GB: Gigabyte</p> <p>INR: Indian Rupees</p> <p>MB: Megabyte</p>			


Table 9-4: Vodafone price plans 25th January 2015, Source: GSMOutlook (2015d).

Key Characteristics of Case Studies

Key characteristics of case studies	
#	Characteristic
1	Case studies are useful, where the ‘case’ is the focus of interest.
2	Case studies may be utilised in both qualitative and quantitative research.
3	The phenomenon to be studied ought to be examined in a natural setting.
4	Data may be collected by multiple means.
5	A case study should limit the examination to one or few actors.
6	These actors are studied intensively in-depth from different perspectives. This includes the linkage to existing theory.
6	Case studies are more suitable for explorations, classifications and hypothesis generation and testing. Therefore, the researcher should have an attitude towards exploring phenomena.
8	No manipulation or experiment controls are involved.
9	Dependent and Independent Variables shall not be specified upfront.
10	Case study results rely on the interpretive and integrative abilities of the researcher and might leave room for a researcher’s interpretation.
11	Changes in the case study research methods could take place to develop new hypothesis.
12	Case studies seem most useful in analysing ‘Why?’ and ‘How?’ questions.
13	The results of case studies are not easily generalisable but good science is problem driven not methodology driven.

Table 9-5: Key characteristics of case studies, Source: Stake (1995), Yin (2003), Flyvbjerg (2006) and Bryman (2012).

Data Collection Instructions

Device Name		SIM / Cell Number	
	Instructions for the Traceroute Data Collection Using the Portolan Network Sensing Architecture Android Application		
Step Nr.	Description	Checkbox	
Step 1	<ol style="list-style-type: none"> 1. Make sure your device is fully charged before beginning the data collection. 2. Note the device name and SIM / cell nr. into the upper section of this instruction. 3. Shut-down your device. 4. Insert the given / acquired SIM-card. 5. Reboot / Restart your device. 		
Step 2	<ol style="list-style-type: none"> 1. Enter the Google Play Store 2. Download Advanced Task Killer* App 3. Download Portolan Network Tools** App <p>* requires >= Android 1.6 **requires >= Android 4.1</p>		
Step 3	<ol style="list-style-type: none"> 1. Stop the Wi-Fi and Bluetooth connections in the settings of your Android device. 2. Open the Advanced Task Killer App and stop / pause all running applications except the ones being named "Google..." 		
Step 4	<ol style="list-style-type: none"> 1. Open the Portolan Network Tools App. 2. Read and Agree to the upcoming Disclaimer (see Figure 1 below). 3. Open the Side-Menu (see Figure 2 below) and select / start the "RSSI – Tracker". 4. The Portolan Network tools App is now automatically measuring a) the data packet traceroutes and b) the received signal strength of your device. 5. Carry the device with you and note-down your data collection journey. 		

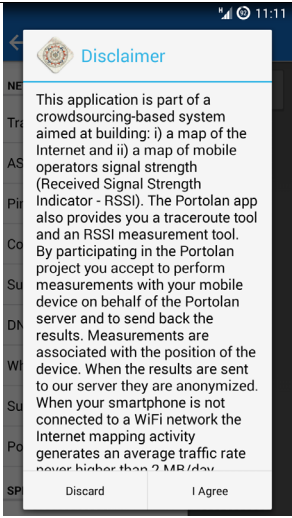


Figure 1: Disclaimer

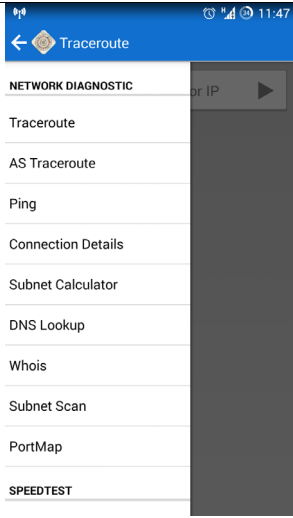


Figure 2: Side-Menu

Notes on the data collection journey:

Please contact your data collection supervisor or Sebastian Sigloch (PhD Student ARU) via sebsigloch@icloud.com or +41799276112 for any question related to this data collection.

Ethical Considerations, Reliability, Validity and Generalisability

Ethical Considerations

The Ethical approval for this dissertation is obtained from the Research Ethics Sub-Committee at Anglia Ruskin University. The primary *traceroute* data collection took place outside the European Economic Area (EEA) and complied with the laws of India. The collected *traceroute* data, obtained in Tamil Nadu, India, was transferred back to Italy in the European Economic Area (EEA), where it was cached and subsequently transferred via email to the data collector in the United Kingdom. The collected primary data neither incorporated any personal information, nor specifically traceable end-user data, since all *traceroute* measurements were collecting randomly chosen (from the Portolan (2015) server) network connection destinations. This data collection complies with both the rules of the European Economic Area (EEA) and the UK Data Protection Act from 1998.

Reliability

The different steps of our warranted assertion inquiry always tested the collected *traceroute* hop observations from the same upstream Internet market structural perspective, representing the primary interest in our single case study design. Hence, our strategy showed a consistent internal reliability. Furthermore, all steps of the *Complex* and *Statistical Network Analysis* were double-checked, demonstrating test-retest reliability. Since our network measurements assessed the construct that we intended to study, our research shows face validity. By evaluating the outcomes of our case study, we provided in-depth information on how well individual Complex Network measurements performed when studying the upstream Internet market structures of the three Tamil Nadu mobile broadband operators. Therefore, our research also shows formative validity. Finally, since all steps of our warranted assertion inquiry can also be repeated by other researchers at any time as long as the same datasets and analytical steps as described are used, we consider our research to be reliable.

Pragmatic Validity

By following a pragmatic paradigm rather than any traditional (post-) positivists or constructivist's philosophy, we encountered pragmatic validity (Worren, Moore and Elliott, 2002). When explanatory scientists drive experiments in controlled laboratories, variables can be minimized and evaluated against internal validity. When looking at

exploratory and pragmatic research, some researchers such as Collins, Joseph and Bielaczuc (2004) refer to the ‘messy situations’ of real-life, where independent variables (covariates) can neither be minimised, nor entirely accounted for. Since our case study represents the modelled realities of real-world Internet structures at different levels of granularity (Internet Protocol and Autonomous System granularity), the complex nature of real-life intervention means that the effect of some interventions of our abducted *Working Hypotheses* is conclusively linkable to the cause itself. However, as pragmatist researchers, we are looking at causal effects through a different perspective based on *Working Hypotheses*. Therefore, our research aims to bring about real-world problem-solving and valuable research artefacts to be used in practice and policy. According to Nowotny (2003), knowledge is considered socially robust when validated by a multidisciplinary community of practice. Parts of this dissertation's methodology followed the peer-reviewed pilot experiment by Giovannetti and Sigloch (2015), resulting, at least partially, in socially robust knowledge.

Generalisability

Case study designs are often referred to as having a low generalisability, especially when working with a single case strategy. However, according to Ragin (1992) and Flyvbjerg (2006), it greatly depends upon the strategic choice of the case one is speaking of. Our research aims and objectives clarified the need for a deeper understanding of the given problems and the importance of understanding their consequences in a real-life context. Therefore, our strategic choice of case study design and the careful selection of the case study object and its subjects were chosen in order to better understand the upstream Internet market structures present in a lower-middle income country such as India. The political scientist Eckstein (2000) asserted that case studies are ‘valuable at all stages of the theory-building process, but most valuable at that stage of theory building where least value is generally attached to them: the stage at which candidate theories are tested’. Therefore, considering our aim to achieve an in-depth insight into the given phenomena of structural bottlenecks in the upstream Internet market structure, our pragmatic validity and our strategic choice for a single case selection with three subjects and different exploratory-quantitative research methods, provides a great transferability but no generalisability to the wider population, see also section 7.4.

Number of observations by originating Autonomous System

Number of observations by originating AS			
#	Organisational Name, Source: Hurricane Electric (2016), (Autonomous System Number)	Number of observations in the primary collected traceroute hop observations	Percentage of observations commencing from originating Autonomous System.
1	Citycom Networks Pvt. Ltd., (AS10029)	556,043	76.04%
2	C48 Okhla Industrial Estate, Vodafone India Ltd., (AS55410)	51,985	7.11%
3	Vodafone Essar Ltd. Telecommunication, (38266)	30,633	4.19%
4	Tata Communications (formerly VSNL), (AS4755)	27,781	3.80%
5	Bharti Airtel Ltd. Telemedia Services, (AS24560)	21,707	2.97%
6	BSNL (Bharat Sanchar Nigam Ltd.), (AS9829)	20,037	2.74%
7	Broadband Pacenet Pvt. Ltd. (AS23682)	10,567	1.45%
8	Aircel Ltd., (AS55831)	4,749	0.65%
9	Idea Cellular Ltd., (AS45271)	3,862	0.53%
10	PT Quasar Jaringan Mandiri, (AS56247)	2,521	0.34%
11	Bharti Airtel Ltd., (AS45609)	956	0.13%
12	Reliance Communications Ltd., (AS18101)	359	0.05%
	Total	731,200	100%
Key			

AS: Autonomous System.

Table 9-6: Number of observations by originating Autonomous System.

Power-law degree distributions using R

Aircel power-law degree distribution using R (2016)
<pre> #Assign library library("powerLaw") #read distribution from text file data<-read.table(file="Dissertation/Aircel_degree_distribution.txt", header=(TRUE) head(data) #define plaw variable plaw <- conpl\$new(data\$Degree) #define est variable est <- estimate_xmin (plaw) #generate power-law distribution plaw\$setXmin(est) #plot power-law distribution plot(plaw, sub="Strength Distribution", xlab="k", ylab="P(k)") legend (x=35, y=0.5, c("degree observations"), col=c("black", "red"), pch=c(1,3)) lines (plaw, col=2, lwd=3) lines (plaw, col="red", lwd=2) #copy distribution as png file and exit dev.copy(png, "Aircel_degree_distribution.png") dev.off() </pre>

Table 9-7: Aircel degree distribution using R (2016).

k-core decomposition using R

Aircel k-core decomposition using R (2016)

```

#Assign igraph library
library(igraph)

#Read csv file without headers, (note: file-path of the *.csv file might vary)
data=read.csv(file="/Dissertation/Step3/Aircel/k-core
decomposition/AS55831_Aircel_for_R.csv",header=FALSE)

#generate matrix from data
matrix=as.matrix(data)

#assign the edgelist to the variable network.
network=graph.edgelist(matrix, directed = TRUE)

#define coreness variable
coreness(network, mode = c("all", "out", "in"))

#define coreness layout and run algorithm
CorenessLayout <- function(g) {
  coreness <- graph.coreness(g);
  xy <- array(NA, dim=c(length(coreness), 2));
  shells <- sort(unique(coreness));
  for(shell in shells) {
    v <- 1 - ((shell-1) / max(shells));
    nodes_in_shell <- sum(coreness==shell);
    angles <- seq(0,360,(360/nodes_in_shell));
    angles <- angles[-length(angles)]; # remove last element
    xy[coreness==shell, 1] <- sin(angles) * v;
    xy[coreness==shell, 2] <- cos(angles) * v;
  }
  return(xy);
}

#assign the coreness variable with the coreness of the imported network
coreness <- graph.coreness(network);

#assign colouring scheme
colbar <- rainbow(max(coreness));

#define kcoredecomposition to be the network layout
kcoredecomposition <- CorenessLayout(network);

#plot the graph, (note: file-path of the plot might vary)

```

```
plot(network,layout=kcoredecomposition,vertex.size=5,vertex.color=colbar[coreness]  
, vertex.label = ifelse(coreness > 10, V(network)$name, NA))  
dev.copy(png, "Aircel_kcore_decomposition.png")  
dev.off()
```

Table 9-8: Aircel k-core decomposition using R (2016).

AS Numbers and Organisational Names

AS Numbers and Organisational Names for originating Autonomous System Numbers (including Wi-Fi originating connections)		
#	Autonomous System Number	Organisational Name of the Autonomous System (Maxmind, 2015)
1	AS10029	Citycom Networks Pvt. Ltd.
2	AS18101	Reliance Communications Ltd.
3	AS23682	Broadband Pacenet Pvt. Ltd.
4	AS24560	Bharti Airtel Ltd. Telemedia Services
5	AS38266	Vodafone Essar Ltd. Telecommunication
6	AS45271	Idea Cellular Ltd.
7	AS45609	Bharti Airtel Ltd. AS for GPRS Service
8	AS4755	TATA Communications formerly VSNL
9	AS55410	C48 Okhla Industrial Estate, Vodafone India Ltd.
10	AS55831	Aircel Ltd.
11	AS56247	PT Quasar Jaringan Mandiri
12	AS9829	BSNL National Internet Backbone
Key AS: Autonomous System		

Table 9-9: Autonomous System Numbers and Organisational Names, Source: Maxmind (2015).

Autonomous System Route Propagations

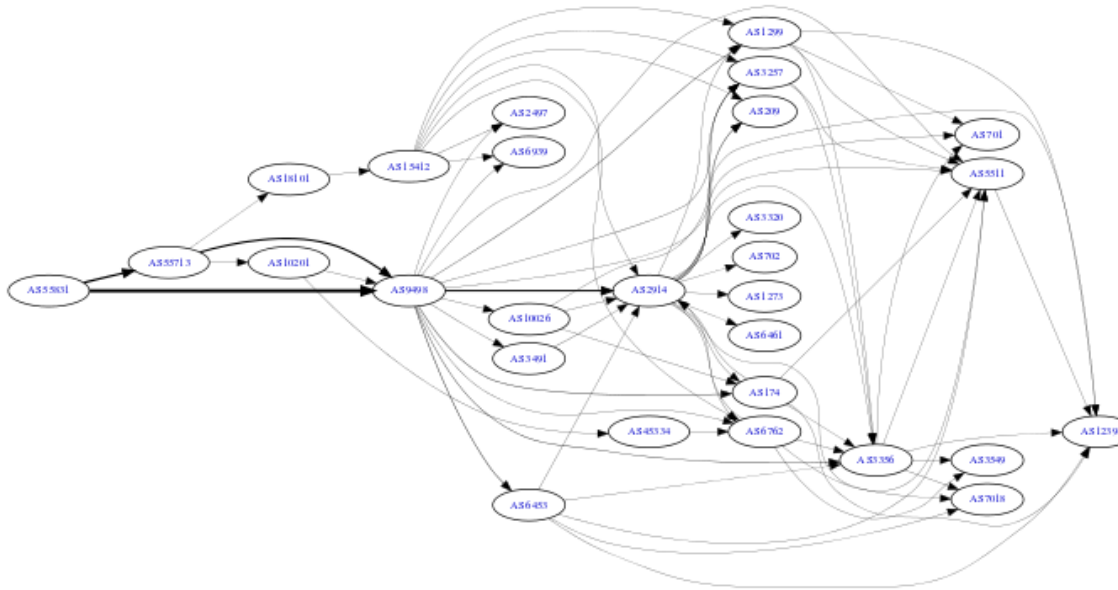


Figure 9-1: Aircel AS55831 Route Propagation, Source: Hurricane Electric (2016).

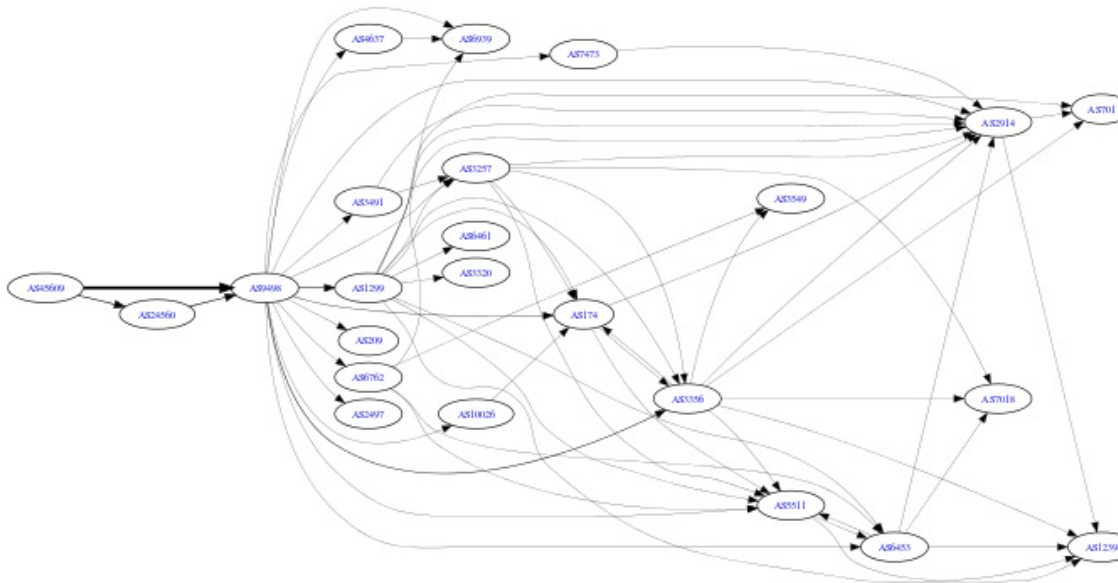


Figure 9-2: Bharti Airtel AS45609 Route Propagation, Source: Hurricane Electric (2016).



Figure 9-3: Vodafone AS38266 Route Propagation, Source: Hurricane Electric (2016).

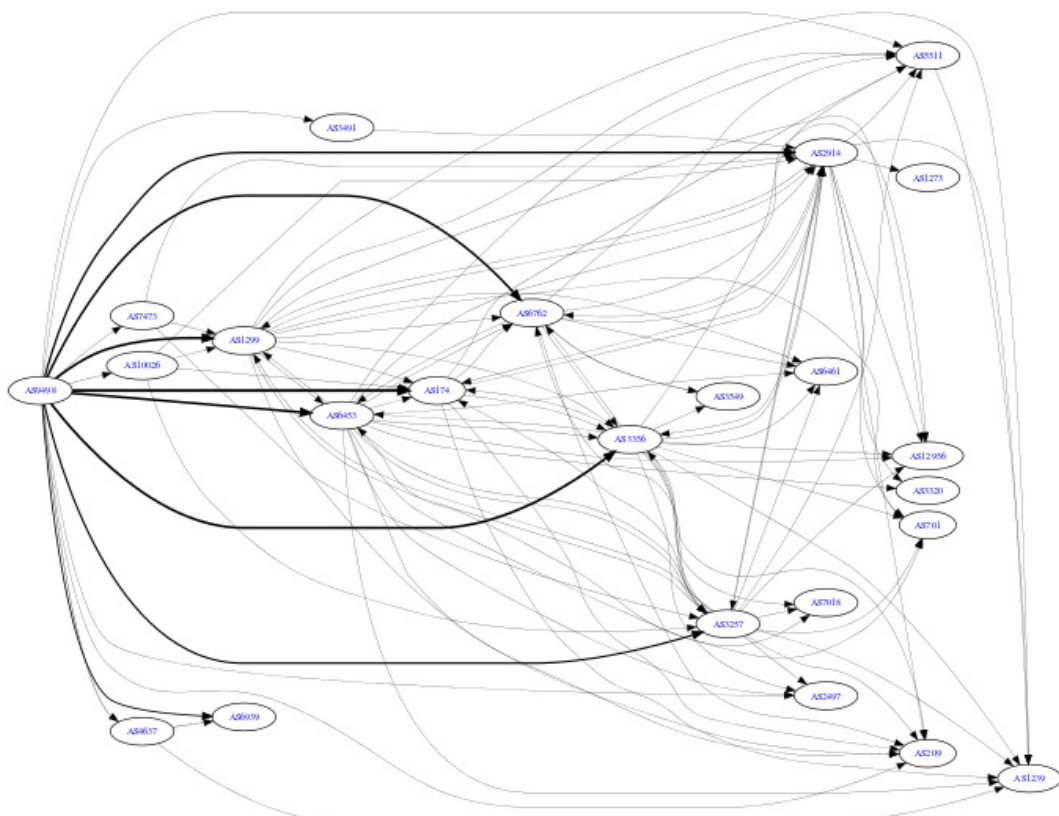


Figure 9-4: Bharti Airtel AS9498 Route Propagation, Source: Hurricane Electric (2016).

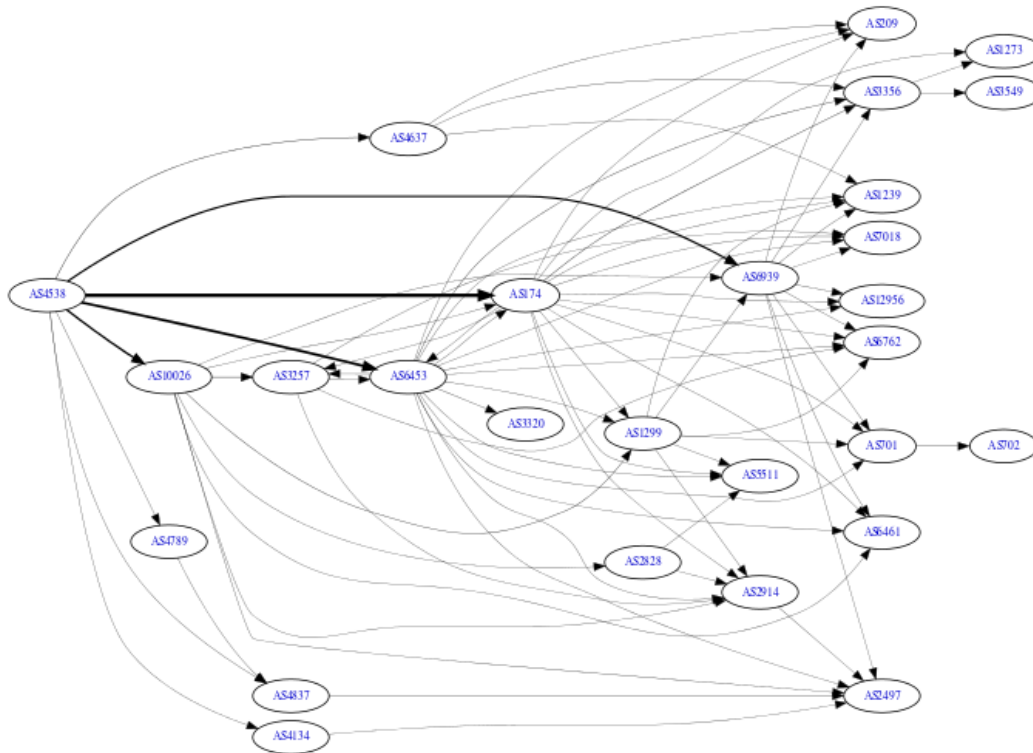


Figure 9-5: China Education and Research Network Center AS4538 Route Propagation, Source: Hurricane Electric (2016).

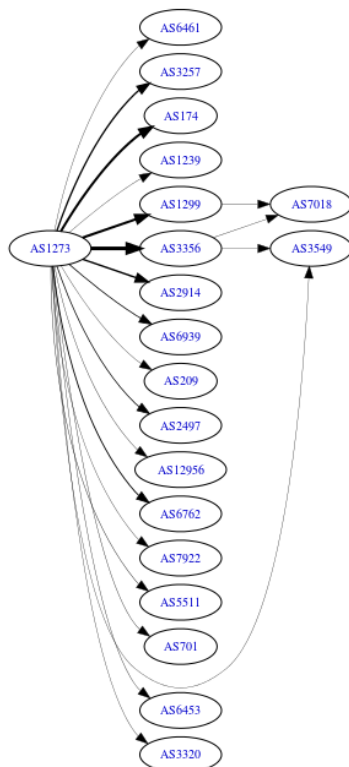


Figure 9-6: Cable and Wireless Worldwide plc. AS1273 Route Propagation, Source: Hurricane Electric (2016).

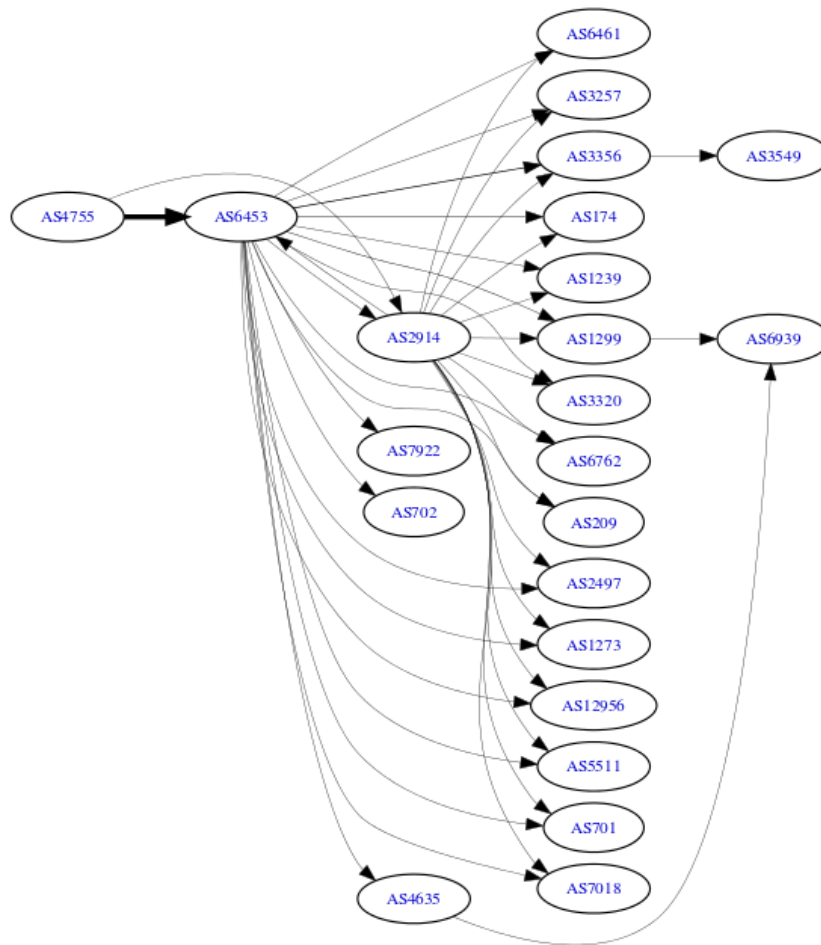


Figure 9-7: Tata Communications (formerly VSNL) AS4755 Route Propagation, Source: Hurricane Electric (2016).

Stata Do-Files

Stata do-file Model 1 and Model 2

Stata do-file Model 1 and Model 2
<pre> *Load Dataset (file location might change) use "/Users/sigloch/Thesis_Files/Inquiry Iteration 8 - Statistical Testing of Working Hypotheses/base_data/data.dta" /* START DESCRIPTIVE STATISTICS */ *Descriptive statistics for the variables summarize woutd winde clus eig *Detailed descriptive statistics for the variables summarize woutd winde clus eig, detail * => Data shows skew and non-normal distributions for woutd and eig. *generate histograms for dependent and independent variables histogram woutd, frequency normal title (Weighted Out-Degree distribution) histogram winde, frequency normal title (Weighted In-Degree distribution) *Two-way scatter-plot and correlation Weighted Out-Degree and Weighted In- Degree scatter winde woutd, title(Scatter plot winde woutd) subtitle(Total observations) corr winde woutd *Operator based two-way scatter-plots Weighted Out-Degree and Weighted In- Degree scatter winde woutd if prov==1, title(Scatter plot winde woutd) subtitle(Aircel) corr winde woutd if prov==1 scatter winde woutd if prov==2, title(Scatter plot winde woutd) subtitle(Bharti Airtel) corr winde woutd if prov==2 scatter winde woutd if prov==3, title(Scatter plot winde woutd) subtitle(Vodafone) corr winde woutd if prov==3 *Histogram of Clustering Coefficient and Eigenvector Centrality. histogram clus, frequency normal title (Clustering Coefficient distribution) </pre>


```
histogram eige, frequency normal title (Eigenvector Centrality distribution)
```

```
*generate ln-transformed variables
```

```
gen lwoutd = ln(woutd)
```

```
gen lwinde = ln(winde)
```

```
gen leige = ln(eige)
```

```
gen lclus = ln(clus)
```

```
*Descriptive statistics for the ln-transformed variables
```

```
summarize lwoutd lwinde lclus leige
```

```
*Detailed descriptive statistics for the ln-transformed variables
```

```
summarize lwoutd lwinde lclus leige, detail
```

```
*Descriptive statistics Aircel
```

```
summarize lwoutd lwinde lclus leige if prov == 1
```

```
*Descriptive statistics Bharti Airtel
```

```
summarize lwoutd lwinde lclus leige if prov == 2
```

```
*Descriptive statistics Vodafone
```

```
summarize lwoutd lwinde lclus leige if prov == 3
```

```
*Operator based two-way scatter-plots Ln(Weighted Out-Degree )and Ln(Weighted In-Degree)
```

```
twoway (scatter lwinde lwoutd) (lfit lwinde lwoutd), title (Twoway lwinde lwoutd with lfit)
```

```
corr lwinde lwoutd
```

```
twoway (scatter lwinde lwoutd if prov==1) (lfit lwinde lwoutd), title (Aircel Twoway lwinde lwoutd with lfit)
```

```
corr lwinde lwoutd if prov==1
```

```
twoway (scatter lwinde lwoutd if prov==2) (lfit lwinde lwoutd), title (Bharti Airtel Twoway lwinde lwoutd with lfit)
```

```
corr lwinde lwoutd if prov==2
```

```
twoway (scatter lwinde lwoutd if prov==3) (lfit lwinde lwoutd), title (Vodafone Twoway lwinde lwoutd with lfit)
```

```
corr lwinde lwoutd if prov==3
```

```

*generate histograms for ln-transformed dependent and independent variables
histogram lwoutd, frequency normal title (Ln(Weighted Out-Degree) distribution)
histogram lwinde, frequency normal title (Ln(Weighted In-Degree) distribution)
histogram lclus, frequency normal title (Ln(Clustering Coefficient) distribution)
histogram leige, frequency normal title (Ln(Eigenvector Centrality) distribution)

/* START FUNCTIONAL FORM SPECIFICATION CHECK MODEL 1 */

*Lin-Lin Regression without displaying results
regress woutd clus eige

*Lin-Log Regression
regress woutd lclus leige

*Log-Lin Regression
regress lwoutd clus eige

*Log-Log Regression
regress lwoutd lclus leige

*Test Linear Variable Relationships Model 1
twoway (scatter lclus lwoutd) (lfit lclus lwoutd), title (Twoway lclus lwoutd with lfit)
corr lclus lwoutd
twoway (scatter leige lwoutd) (lfit leige lwoutd), title (Twoway leige lwoutd with lfit)
corr leige lwoutd

twoway (scatter lclus lwoutd) (lfit lclus lwoutd), by(prov) title (Twoway lclus lwoutd
with lfit by prov)
twoway (scatter leige lwoutd) (lfit leige lwoutd), by(prov) title (Twoway leige lwoutd
with lfit by prov)

/* START FUNCTIONAL FORM SPECIFICATION CHECK MODEL 2 */

*Lin-Lin Regression without displaying results
regress winde clus eige

```

*Lin-Log Regression

regress winde lclus leige

*Log-Lin Regression

regress lwinde clus eige

*Log-Log Regression

regress lwinde lclus leige

*Test Linear Variable Relationships Model 2

twoway (scatter lclus lwinde) (lfit lclus lwinde), title (Twoway lclus lwinde with lfit)
corr lclus lwinde

twoway (scatter leige lwinde) (lfit leige lwinde), title (Twoway leige lwinde with lfit)
corr leige lwinde

twoway (scatter lclus lwinde) (lfit lclus lwinde), by(prov) title (Twoway lclus lwinde
with lfit by prov)

twoway (scatter leige lwinde) (lfit leige lwinde), by(prov) title (Twoway leige lwinde
with lfit by prov)

/* ECONOMETRIC MODEL 1 Ln(Weighted Out-Degree) BY PROVIDER */

/* AIRCEL */

*Normal Regression for Aircel

regress lwoutd lclus leige if prov == 1

*Heteroskedasticity test

hettest

*Robust Regression for Aircel

regress lwoutd lclus leige if prov == 1, vce(robust)

*Ramsey RESET test

ovtest

*Testing Variables due to omitted variable bias

```
regress lclus leige if prov == 1
```

```
regress lwoutd lclus if prov == 1
```

```
regress lwoutd leige if prov == 1
```

*Generate independent variables at 2nd power to check omitted variables

```
gen lclus2 = lclus^2
```

```
gen leige2 = leige^2
```

*Test Model at 2nd power

```
regress lwoutd lclus leige lclus2 leige2 if prov == 1, vce(robust)
```

*Ramsey RESET test

```
ovtest
```

*Generate independent variables at 3rd power to check omitted variables

```
gen lclus3 = lclus^3
```

```
gen leige3 = leige^3
```

*Test Model at 3rd power

```
regress lwoutd lclus leige lclus2 leige2 lclus3 leige3 if prov == 1, vce(robust)
```

*Ramsey RESET test

```
ovtest
```

*Drop generated variables of 2nd and 3rd power.

```
drop lclus2
```

```
drop leige2
```

```
drop lclus3
```

```
drop leige3
```

*Quietly rerun initial regression

```
quietly regress lwoutd lclus leige if prov == 1, vce(robust)
```

*Variance Inflation Factors Test

```

vif

*Calculate the error term for the residual
predict saircel, residual

*Generate Histogram for the residual distribution
histogram saircel, frequency normal title (Aircel residual distribution Model 1) note
("Source: Elaborated by the author of this dissertation in using Stata (2016).")

/* BHARTI AIRTEL */

*Normal Regression for Bharti Airtel
regress lwoutd lclus leige if prov == 2

*Heteroskedasticity test
hettest

*Robust Regression for Bharti Airtel
regress lwoutd lclus leige if prov == 2, vce(robust)

*Ramsey RESET test
ovtest

*Testing Variables due to omitted variable bias
regress lclus leige if prov == 2
regress lwoutd lclus if prov == 2
regress lwoutd leige if prov == 2

*Generate independent variables at 2nd power to check omitted variables
gen lclus2 = lclus^2
gen leige2 = leige^2

*Test Model at 2nd power
regress lwoutd lclus leige lclus2 leige2 if prov == 2, vce(robust)

*Ramsey RESET test
ovtest

```

```

*Generate independent variables at 3rd power to check omitted variables
gen lclus3 = lclus^3
gen leige3 = leige^3

*Test Model at 3rd power
regress lwoutd lclus leige lclus2 leige2 lclus3 leige3 if prov == 2, vce(robust)

*Ramsey RESET test
ovtest

*Drop generated variables of 2nd and 3rd power.
drop lclus2
drop leige2
drop lclus3
drop leige3

*Quietly rerun initial regression
quietly regress lwoutd lclus leige if prov == 2, vce(robust)

*Variance Inflation Factors Test
vif

*Calculate the error term for the residual
predict sbhartiartel, residual
*Generate Histogram for the residual distribution
histogram sbhartiartel, frequency normal title (Bharti Airtel residual distribution
Model 1) note ("Source: Elaborated by the author of this dissertation in using Stata
(2016).")

/* Vodafone */
*Normal Regression for Bharti Airtel
regress lwoutd lclus leige if prov == 3

*Heteroskedasticity test
hettest

```

```

*Robust Regression for Bharti Airtel
regress lwoutd lclus leige if prov == 3, vce(robust)

*Ramsey RESET test
ovtest

*Variance Inflation Factors Test
vif

*Calculate the error term for the residual
predict svodafone, residual
*Generate Histogram for the residual distribution
histogram svodafone, frequency normal title (Vodafone residual distribution Model 1)
note ("Source: Elaborated by the author of this dissertation in using Stata (2016).")

*Drop generated variables for residual distribution
drop saircel
drop sbhartiartel
drop svodafone

/* ECONOMETRIC MODEL 2 (Weighted In-Degree) BY PROVIDER */

/* AIRCEL */
*Normal Regression for Aircel
regress lwinde lclus leige if prov == 1

*Heteroskedasticity test
hettest

*Robust Regression for Aircel
regress lwinde lclus leige if prov == 1, vce(robust)

*Ramsey RESET test
ovtest

*Testing Variables due to omitted variable bias

```

```

regress lclus leige if prov == 1
regress lwinde lclus if prov == 1
regress lwinde leige if prov == 1

*Independent variables at 2nd power to check omitted variables already created
gen lclus2 = lclus^2
gen leige2 = leige^2

*Test Model at 2nd power
regress lwinde lclus leige lclus2 leige2 if prov == 1, vce(robust)

*Ramsey RESET test
ovtest

*Generate independent variables at 3rd power to check omitted variables
gen lclus3 = lclus^3
gen leige3 = leige^3

*Test Model at 3rd power
regress lwinde lclus leige lclus2 leige2 lclus3 leige3 if prov == 1, vce(robust)

*Ramsey RESET test
ovtest

*Drop generated variables of 2nd and 3rd power.
drop lclus2
drop leige2
drop lclus3
drop leige3

*Quietly rerun initial regression
quietly regress lwinde lclus leige if prov == 1, vce(robust)

*Variance Inflation Factors Test
vif

```



```

*Calculate the error term for the residual
predict saircel, residual

*Generate Histogram for the residual distribution
histogram saircel, frequency normal title (Aircel residual distribution Model 2) note
("Source: Elaborated by the author of this dissertation in using Stata (2016).")

/* BHARTI AIRTEL */

*Normal Regression for Bharti Airtel
regress lwinde lclus leige if prov == 2

*Heteroskedasticity test
hettest

*Robust Regression for Bharti Airtel
regress lwinde lclus leige if prov == 2, vce(robust)

*Ramsey RESET test
ovtest

* => omitted variable bias

*Testing Variables due to omitted variable bias
regress lclus leige if prov == 2
regress lwinde lclus if prov == 2
regress lwinde leige if prov == 2

*Generate independent variables at 2nd power to check omitted variables
gen lclus2 = lclus^2
gen leige2 = leige^2

*Test Model at 2nd power
regress lwinde lclus leige lclus2 leige2 if prov == 2, vce(robust)

*Ramsey RESET test
ovtest

*Generate independent variables at 3rd power to check omitted variables

```

```

gen lclus3 = lclus^3
gen leige3 = leige^3

*Test Model at 3rd power
regress lwinde lclus leige lclus2 leige2 lclus3 leige3 if prov == 2, vce(robust)

*Ramsey RESET test
ovtest

*Drop generated variables of 2nd and 3rd power.
drop lclus2
drop leige2
drop lclus3
drop leige3

*Quietly rerun initial regression
quietly regress lwinde lclus leige if prov == 2, vce(robust)

*Variance Inflation Factors Test
vif

*Calculate the error term for the residual
predict sbhartiairtel, residual
*Generate Histogram for the residual distribution
histogram sbhartiairtel, frequency normal title (Bharti Airtel residual distribution
Model 2) note ("Source: Elaborated by the author of this dissertation in using Stata
(2016).")

/* Vodafone */
*Normal Regression for Bharti Airtel
regress lwoutd lclus leige if prov == 3

*Heteroskedasticity test
hettest

*Robust Regression for Bharti Airtel

```

```
regress lwinde lclus leige if prov == 3, vce(robust)
```

```
*Ramsey RESET test
```

```
ovtest
```

```
*Testing Variables due to omitted variable bias
```

```
regress lclus leige if prov == 3
```

```
regress lwinde lclus if prov == 3
```

```
regress lwinde leige if prov == 3
```

```
*Variance Inflation Factors Test
```

```
vif
```

```
*Calculate the error term for the residual
```

```
predict svodafone, residual
```

```
*Generate Histogram for the residual distribution
```

```
histogram svodafone, frequency normal title (Vodafone residual distribution Model 2)  
note ("Source: Elaborated by the author of this dissertation in using Stata (2016).")
```

```
*Drop generated variables for residual distribution
```

```
drop saircel
```

```
drop sbhartiartel
```

```
drop svodafone
```

Stata do-file Model 3.1 and Model 3.2

Stata do-file Model 3.1 and Model 3.2

```

*Import Data Model 3
import delimited /Users/sigloch/Desktop/Model3.csv

*DESCRIPTIVE STATISTICS*

*Descriptive statistics for the variables
summarize

*Descriptive statistics Aircel
summarize if prov == 1

*Descriptive statistics Bharti Airtel
summarize if prov == 2

*Descriptive statistics Vodafone
summarize if prov == 3

*Scatter Diagram Price Datainmb
twoway (scatter price datainmb) (lfit price datainmb), by(prov) title (Twoway price
datainmb with lfit by prov)

*Scatter Diagram d_lclus priceper
twoway (scatter lclus_hat pricepermb) (lfit lclus_hat pricepermb), by(prov) title
(Twoway lclus_hat pricepermb with linear fit by prov)

*Scatter Diagram d_leige priceper
twoway (scatter leige_hat pricepermb) (lfit leige_hat pricepermb), by(prov) title
(Twoway leige_hat pricepermb with linear fit by prov)

*Histogram Price Distribution
histogram price, by(prov) normal title (Price distribution)

*Histogram Data in MB Distribution
histogram datainmb, by(prov) normal title (Data in MB distribution)

*Histogram Price-Performance Distribution

```

histogram pricepermb, by(prov) normal title (Price per MB distribution per provider)

*Histogram Validity in Days Distribution

histogram vin, by(prov) normal title (Validity in Days distribution per provider)

SPECIFICATION OF MODEL 3

SPECIFICATION OF MODEL 3.1

*generate ln-transformed pricepermb

gen lpricepermb = ln(pricepermb)

*Histogram Price-Performance Distribution

histogram lpricepermb, by(prov) normal title (LN(Price per MB distribution per provider))

*Regression Lclus

regress lpricepermb lclus_hat

*Heteroskedasticity test

hettest

*Ramsey RESET test

ovtest

*Calculate the error term for the residual

predict sprice, residual

*Generate Histogram for the residual distribution

histogram sprice, frequency normal title (Residual distribution Model 3.1)

*Drop error term of the residual

drop sprice

SPECIFICATION OF MODEL 3.2

*Regression

regress lpricepermb leige_hat

*Heteroskedasticity test

hettest

*Ramsey RESET test

ovtest

*Calculate the error term for the residual

predict sprice, residual

*Generate Histogram for the residual distribution

histogram sprice, frequency normal title (Residual distribution Model 3.2)

Stata do-file Model 4

Stata do-file Model 4
<p>*Import Data Model 4 import delimited /Users/sigloch/Desktop/Model4.csv</p> <p>*Descriptive statistics for the variables summarize</p> <p>*Descriptive statistics Aircel summarize if prov == 1</p> <p>*Descriptive statistics Bharti Airtel summarize if prov == 2</p> <p>*Descriptive statistics Vodafone summarize if prov == 3</p> <p>*Scatter Diagram Price Datainmb tway (scatter price datainmb) (lfit price datainmb), by(prov) title (Tway price datainmb with lfit by prov)</p> <p>*Scatter Diagram d_lclus priceper tway (scatter c_lclus pricepermb) (lfit c_lclus pricepermb), by(prov) title (Tway c_lclus pricepermb with lfit by prov)</p> <p>*Scatter Diagram d_leige priceper tway (scatter c_leige pricepermb) (lfit c_leige pricepermb), by(prov) title (Tway c_leige pricepermb with lfit by prov)</p> <p>*Histogram Price Distribution histogram price, by(prov) normal title (Price distribution)</p> <p>*Histogram Data in MB Distribution histogram datainmb, by(prov) normal title (Data in MB distribution)</p> <p>*Histogram Price-Performance Distribution histogram pricepermb, by(prov) normal title (Price per MB distribution per provider)</p>

*Histogram Validity in Days Distribution

histogram vin, by(prov) normal title (Validity in Days distribution per provider)

gen lpricepermb = ln(pricepermb)

*Regression

regress lpricepermb lclus_hat leige_hat vin

*Heteroskedasticity test

hettest

*Ramsey RESET test

ovtest

* => omitted variables bias

*Multicollinearity

vif

*Calculate the error term for the residual

predict sprice, residual

*Generate Histogram for the residual distribution

histogram sprice, frequency normal title (Residual distribution Model 4)

Stata do-file Quality of Service Correlation

Correlation do-file
<pre>import delimited "/Users/sigloch/Thesis_Files/Inquiry Iteration 8 - Statistical Testing of Working Hypotheses/data/Model6/Model6.csv", delimiter(comma) varnames(1) case(preserve) *summarize the findings summarize *corr variables corr c_lclus_m1 c_leige_m1 c_lclus_m2 c_leige_m2 c_loweige sprovd trans_d trans_u speed_d throughput latency pdp drop_rate</pre>

Additional Graph Visualisations

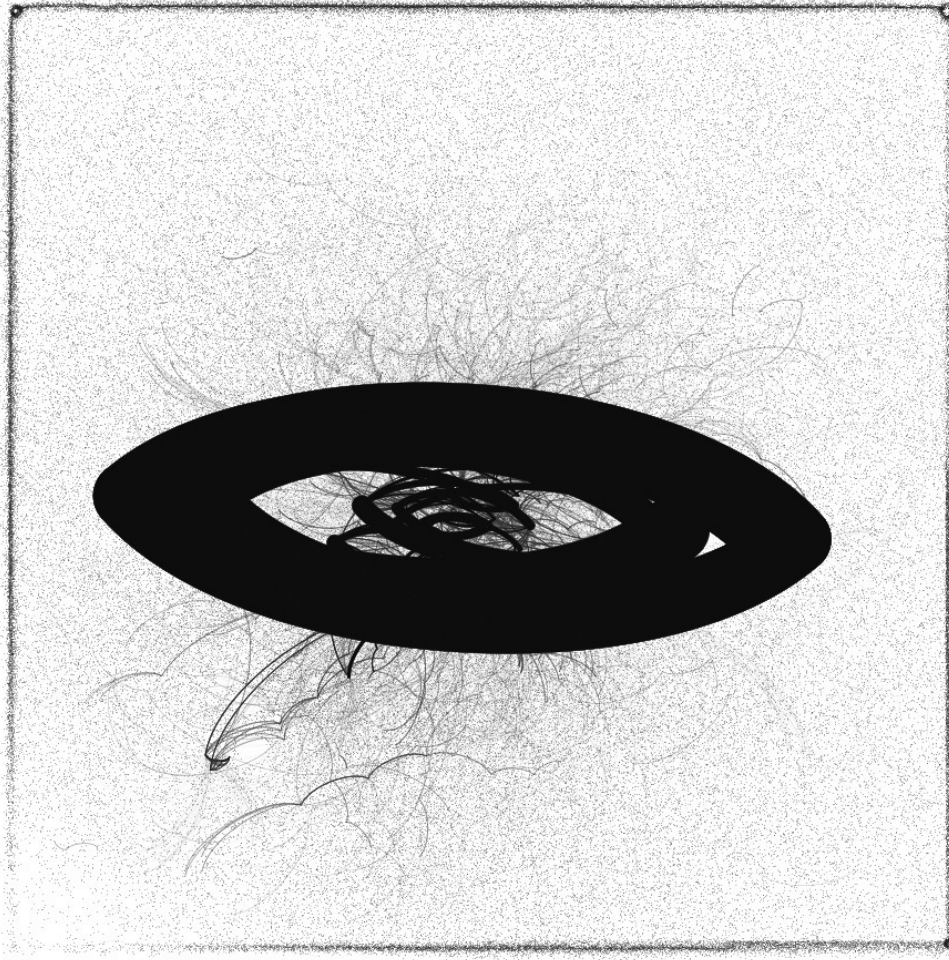


Figure 9-8: 'The Eye' - Total 731,200 observations at IP granularity using the Barábasi-Albert Standard Model with thick edges, elaborated using Gephi (2016).

